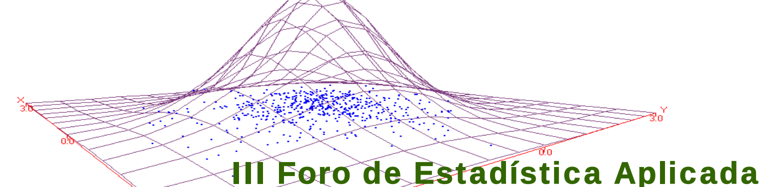


III Foro de Estadística Aplicada

II Congreso Nacional de Estadística

II Meeting of the International Biometric Society, Guatemala Group

Ciudad de Guatemala, 2003.

3 al 7 de noviembre

Minicurso

Análisis estadístico de datos usando Microsoft Excel®

Ing. Agr. M.A. Ezequiel Abraham López Bautista[1]

Ing. Agr. Byron Humberto González Ramírez[2]

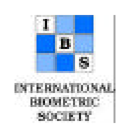
[1] Ingeniero Agrónomo en Sistemas de Producción Agrícola (FAUSAC, Guatemala) Mestre em Agronomía Área de concentración: Estadística y Experimentación Agronómica (ESALQ/USP Brasil). Catedrático de la Subárea de Métodos de Cuantificación e Investigación (FAUSAC)

[2] Ingeniero Agrónomo en Sistemas de Producción Agrícola (FAUSAC, Guatemala). Maestrando en Nuevas Tecnologías de la Información y Comunicación- UNED España- Director del Centro de Telemática, Facultad de Agronomía (USAC)










I. Presentación

La estadística es una ciencia que atraviesa transversalmente la mayoría de estudios en cualquiera de los campos particulares de la ciencia. Decisiones sobre procedimientos o metodologías a aplicarse, son apoyadas por el estudio de la información y su trato estadístico. Inferencias, calibraciones de equipo, ahorro de recursos entre otros, son ejemplos directos del aporte de la estadística a los procesos productivos.

De la misma manera como todas las ciencias se han visto beneficiadas del uso de la informática para acelerar sus procesos, la estadística también se ha visto influenciada por el cada vez más creciente mercado de software estadístico. Sin embargo, aún cuando en la actualidad existen numerosos programas de esta categoría, la mayoría de ellos presentan inconvenientes en su uso, principalmente altos costos y complejidad.

La situación mencionada ha obligado a buscar software de fácil empleo, intuitivo y principalmente cuyo costo no sea elevado. Es aquí donde precisamente Microsoft Excel[*] aparece como una de las mejores alternativas por cuanto disponemos inmediatamente de el al instalar la Suite Office. Adicionalmente la facilidad en el manejo de datos usando planillas de cálculo, hacen de Microsoft Excel el software ideal para simplificar la aplicación de metodologías estadísticas.

El curso planteado integra la aplicación de plantillas de cálculo de Microsoft Excel en las metodologías y técnicas estadísticas enseñadas en los primeros cursos de estadística a nivel superior, aunque también se discuten procedimientos intermedios y algunos avanzados. En todo momento se ha pensado en la disponibilidad que la mayoría de personas tiene de Microsoft Excel, y el soporte de encontrar este software en todas las versiones de Windows e incluso Macintosh.

Contenidos Generales

1. Análisis Exploratorio de Datos
2. Construcción de tablas dinámicas
3. Distribuciones de probabilidad continuas y discretas
4. Construcción de cajas de dispersión
5. Pruebas de medias dependientes e independientes
6. Análisis de Varianza
7. Análisis de Regresión

---

[*]Excel es una marca registrada de Microsoft Corporation.

1. ANÁLISIS EXPLORATORIO DE DATOS

El Análisis Exploratorio de Dados, antiguamente llamado Estadística Descriptiva, constituye lo que la mayoría de las personas entiende como Estadística, e inconscientemente se usa a diario. Consiste en RESUMIR Y ORGANIZAR los datos colectados a través de tablas, gráficos o medidas numéricas, y a partir de los datos resumidos buscar alguna regularidad o patrón en las observaciones (INTERPRETAR los datos).

A partir de esa interpretación inicial es posible identificar se los datos siguen algún modelo conocido, que permita estudiar el fenómeno bajo análisis, o si es necesario sugerir un nuevo modelo.

El objetivo de esta unidad es presentar los principales procedimientos de Análisis Exploratorio de Dados, tal como son presentados en la sala de aula, pero utilizando la planilla electrónica Excel. Los dados están disponibles en el archivo AulaExcel1.xls, disponible en: C:\\Mis Documentos\Estadística\

Y contiene informaciones sobre 474 empleados de una empresa, a través de 9 variables, descritas a continuación:

Sexo = sexo del empleado

Edad = edad del empleado, expresada en años

AñosEd = años de educación

Función = función que ocupa dentro de la empresa

SalarioA = salario anual (expresados en Quetzales)

AñosS = años de servicio

ExperP = experiencia (años)

Los procedimientos fueron preparados utilizando Microsoft Excel XP, aunque son similares a los de otras versiones de esta hoja de cálculo.

1.1 Procedimientos para variables cualitativas

Cuando se desea realizar un análisis exploratorio en que únicamente se toman en consideración variables cualitativas, lo más usual es construir tablas de frecuencia para cada variable individualmente, o tablas de contingencia (clasificación doble) relacionando dos variables. Los gráficos pueden ser generados a partir de las tablas. Para poder trabajar con variables aleatorias cualitativas en Excel necesitamos utilizar las **tablas dinámicas**. Vea los siguientes ejemplos.

1.1.1 Tabla de frecuencia y gráfico de la variable SEXO

a) Busque el menú DATOS en la barra principal de Excel. Luego busque la opción INFORME DE TABLAS Y GRAFICOS DINAMICOS, vea la Figura 1. Seleccione esta opción.

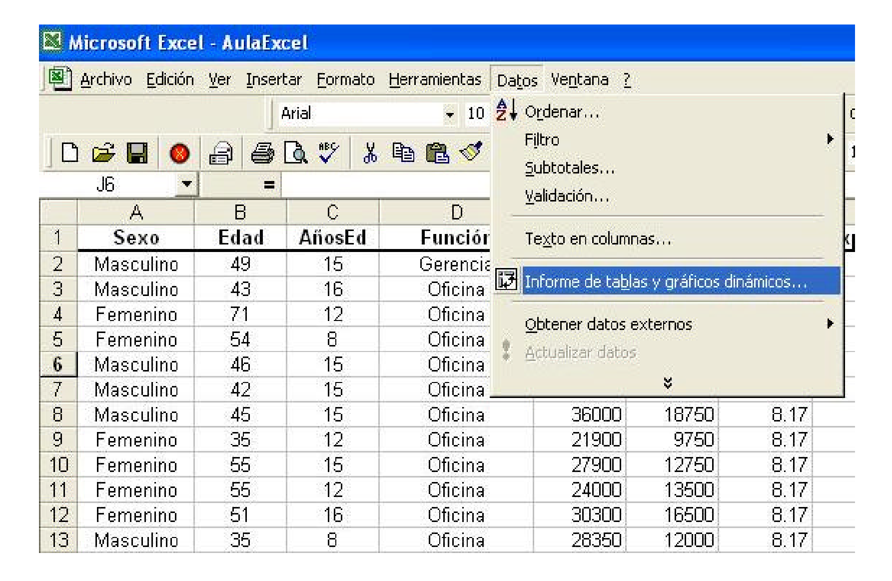


Figura 1 Opción de tabla dinámica

b) Al seleccionar la opción señalada en la Figura 1, surgirá en la pantalla un cuadro similar al mostrado en la Figura 2. Se trata de la etapa 1 de 3 (en Excel 97 se necesitan 4 pasos)

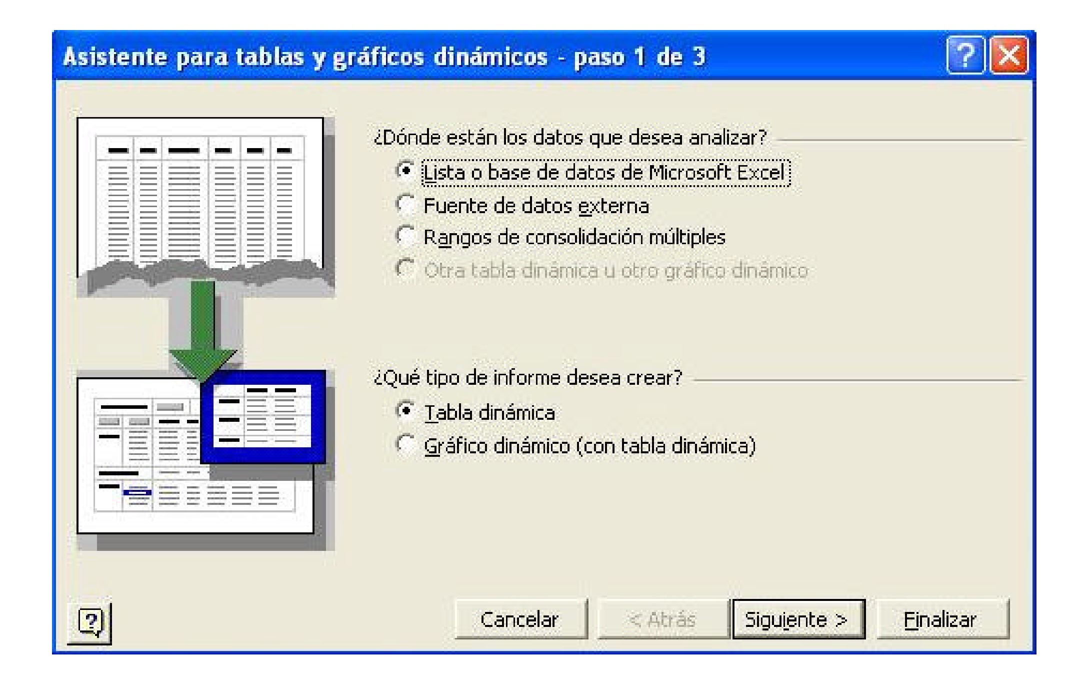


Figura 2 Etapa 1 de la tabla dinámica

Luego es necesario informar dónde están los datos de interés. Usualmente están en la propia planilla de Excel, pueden sin embargo ser apenas una variable o un conjunto de

variables. En este ejemplo, los datos realmente están en una lista de Microsoft Excel, mantenga como está y presione SIGUIENTE, lo que lo llevará a la Figura 3.

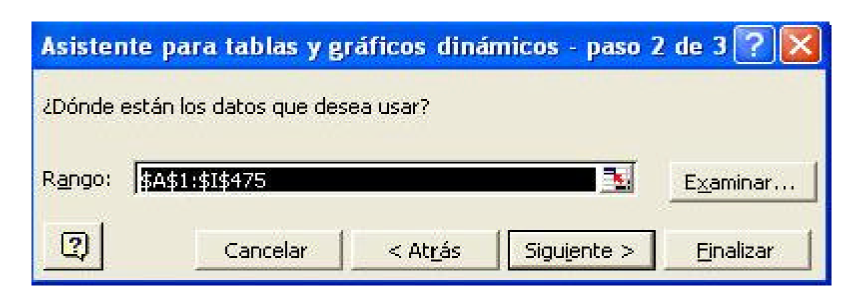


Figura 3 Etapa 2 de la tabla dinámica

c) En este momento es necesario especificar exactamente dónde están los datos de la tabla dinámica: un intervalo de celdas del Excel, indicando las columnas (marcadas con letras) e las filas (marcadas con números). Muchas veces, como en la Figura 3, Excel automáticamente selecciona todos los datos disponibles en la planilla (en el caso de la celda A1 hasta la celda I475, comprendiendo las 474 observaciones de cada una de las nueve variables). Sin embargo, si eso no ocurre, Ud. podrá seleccionar el intervalo por cuenta propia, de dos maneras:

- Escribiendo en el campo RANGO (vea la Figura 3) las referencias de las celdas A1:I475, lo que significa que todas las celdas del intervalo serán consideradas.

- Buscando y seleccionando el intervalo en la planilla; para lo cual, necesita presionar la pequeña flecha roja situada inmediatamente a la derecha del intervalo, lo que resultará en la Figura 4



Figura 4 Cuadro de selección del rango de datos

Con la selección en este formato podrá buscar el intervalo deseado y seleccionarlo utilizando el *mouse*. Para retornar a la situación de la Figura 3 basta presionar la pequeña flecha roja nuevamente. Presionando SIGUIENTE en la pantalla mostrada en la Figura 3 llegamos a la Figura 5.

d) Una vez seleccionados los datos es preciso definir la disposición de la tabla dinámica, cómo estarán distribuidos los datos, y que acción deberá ejecutar la tabla dinámica. La etapa 3 se muestra en la Figura 5.

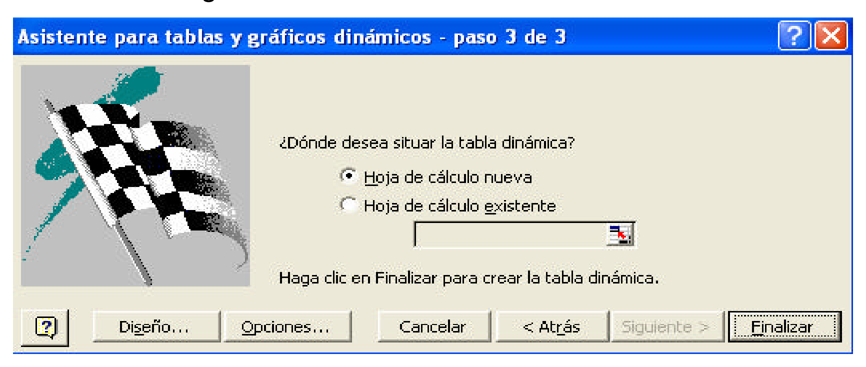


Figura 5 Etapa 3 de la tabla dinámica

Para definir la distribución de los datos, se da un clic en el botón DISEÑO que se muestra en la Figura 6, y se desplegará el siguiente cuadro:



Figura 6 Disposición de la tabla dinámica

Observe a la derecha los nombres de las variables existentes en el archivo de datos, y que fueron seleccionados en la etapa anterior. Como existe interés únicamente en la variable SEXO debemos seleccionarla y arrastrarla hasta el campo FILA, o COLUMNA. Las otras variables no formarán parte de la tabla. Debemos arrastrar también la variable SEXO para el campo DATOS. Eso es necesario para especificar la acción que la tabla deberá ejecutar. La acción a realizar es el conteo de los valores, tal como es mostrado en la Figura 7.



Figura 7 Tabla dinámica para la variable sexo

Si presiona dos veces sobre "CONTAR DE SEXO" surgirá la pantalla expuesta en la Figura 8, donde están las diversas acciones posibles (suma, contar, promedio, máximo, mínimo, desviación estándar, etc., que serán de utilidad cuando se analizarán variables cuantitativas)



Figura 8 Acciones posibles en la tabla dinámica

En el presente caso queremos realmente contar el número de ocurrencias de los valores de la variable SEXO, por tanto, no se efectúan cambios, y se presiona ACEPTAR (Figura 8), y nos lleva a la Figura 5. También se pueden alterar las definiciones de la tabla, presionando el botón "OPCIONES" (Figura 5), que ocasionará la exhibición de la Figura 9.

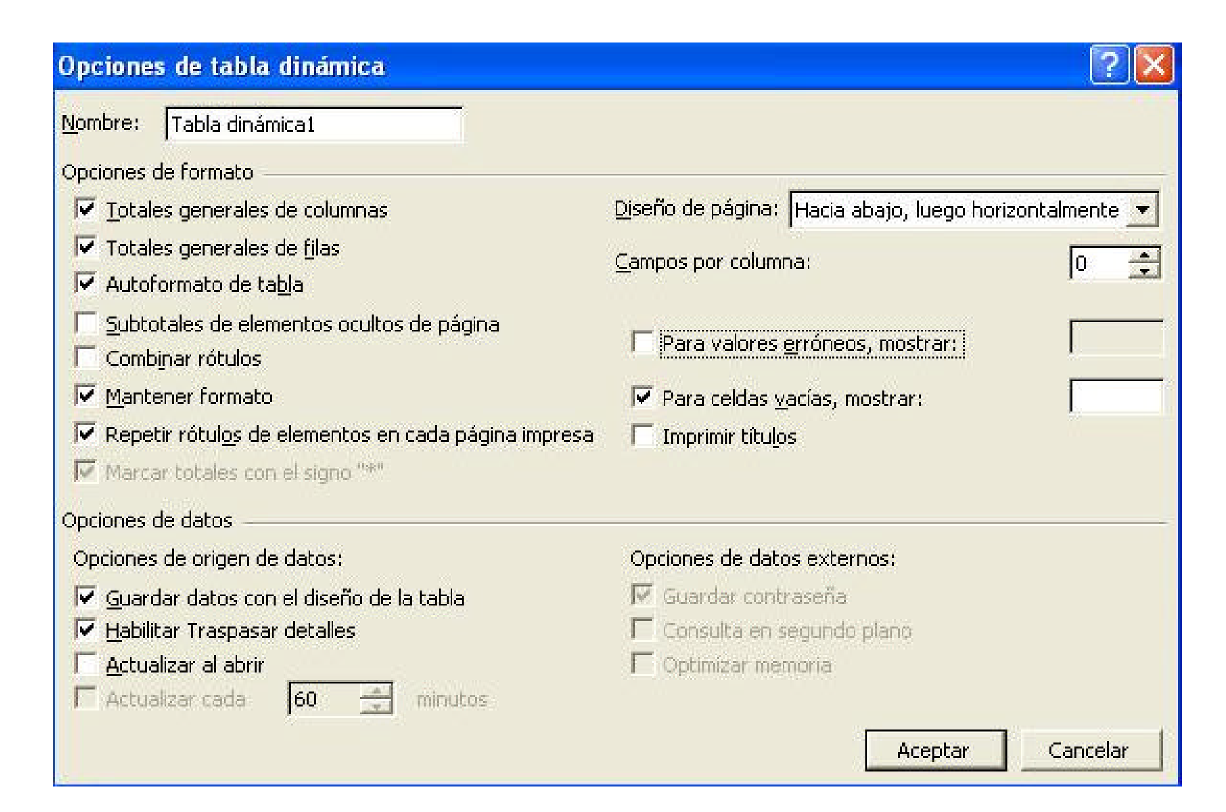


Figura 9 Opciones de presentación de la tabla dinámica

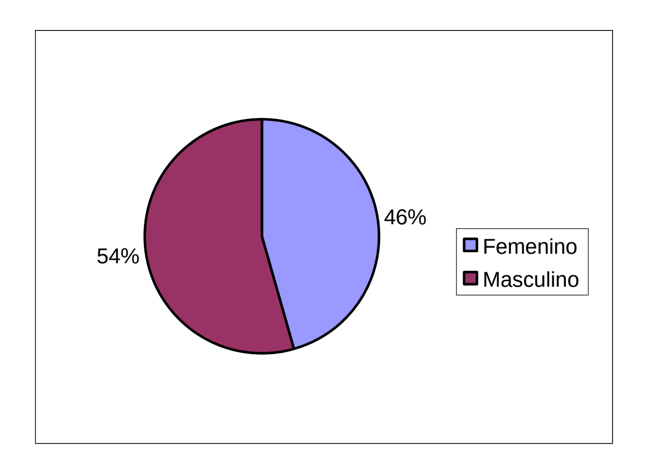
Puede ser que no se deseen que la tabla presente los totales generales para las columnas, o para las filas, dependiendo del caso. Es importante resaltar que los totales de interés pueden ser calculados por Excel, cuando la tabla esté terminada.

Luego de definir donde se colocará la tabla dinámica, y realizadas o no las modificaciones en su presentación, basta con presionar ACEPTAR (Figura 9) y luego FINALIZAR (Figura 5), y el resultado deberá ser:

A	B	C
	Contar de Sexo	
	Sexo	Total
	Femenino	216
	Masculino	258
	Total general	474

Figura 10 Tabla dinámica de la variable sexo

A partir de los resultados de la tabla dinámica, es posible también construir gráficos. Para realizar esta actividad, se busca el icono del "ASISTENTE GRAFICO" y se podrá seleccionar el gráfico más apropiado para los datos (columnas o pizza), pudiendo modificar una serie de aspectos en su apariencia, tales como escala, colores, títulos, entre otras. Los gráficos generados son:



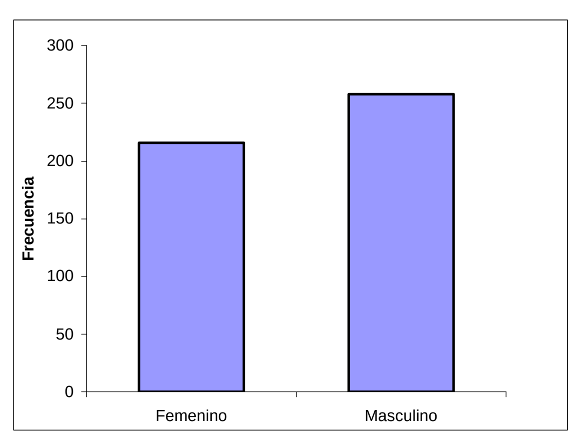


Figura 11 Gráficos generados a partir de la tabla de la variable sexo

1.1.2 Tabla de frecuencias de la variable FUNCION por SEXO

El procedimiento es similar al descrito en el inciso 1.1.1, pero ahora serán utilizadas dos variables, teniendo como propósito construir una tabla de contingencia. Las instrucciones descritas en las literales a) a la d) del inciso 1.1.1 pueden ser repetidas literalmente. Las diferencias comienzan a aparecer cuando se hace el diseño de la tabla. Sí se está construyendo la tabla luego de haber construido la tabla de la variable SEXO, Excel presentará el siguiente aviso:

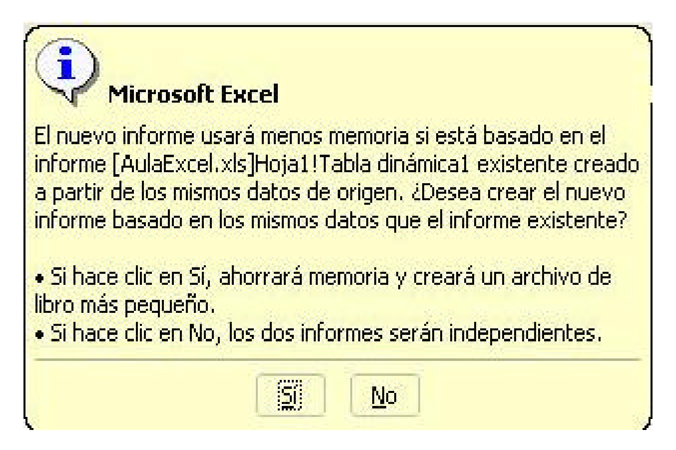


Figura 12 Aviso sobre aprovechamiento de la memoria

Como los datos de origen de la tabla SEXO por FUNCION son los mismos de la tabla SEXO (porque se escogió incorporar todos los datos de la planilla), se puede optar por basar la nueva tabla en la tabla existente, lo que puede ahorrar algo de memoria. Independientemente de la decisión que se tome, es necesario definir la disposición de la tabla.

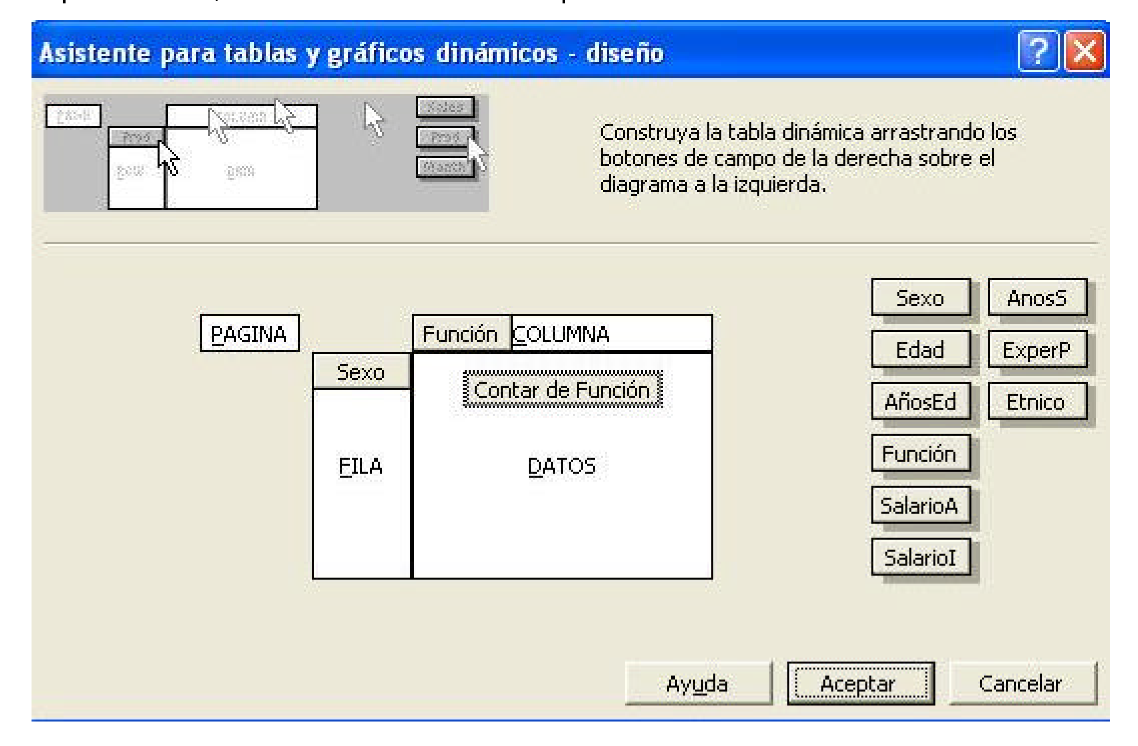


Figura 13 Disposición de la tabla dinámica de SEXO por FUNCION

Seleccione y arrastre la variable SEXO para el campo **FILA**, y la variable FUNCION para el campo **COLUMNA** (o viceversa). Seleccione la variable SEXO (o la variable FUNCION, pero apenas una de ellas) y arrastre para el campo DATOS: la acción estándar es: CONTAR DE FUNCION (o CONTAR DE SEXO, sí se seleccionó esta variable). Presionando Aceptar en la pantalla expuesta en la Figura 13, se llegará a la pantalla expuesta en la Figura 5, donde deberá dar un clic en la opción FINALIZAR, y la tabla resultante será:

Contar de Función	Función			
Sexo	Gerencia	Oficina	Servicios generales	Total general
Femenino	10	206		216
Masculino	74	157	27	258
Total general	84	363	27	474

Figura 14 Tabla de contingencia: Sexo por Función

Para construir gráficos, basta con seleccionar la tabla entera y buscar el icono del asistente gráfico y seleccionar el tipo de gráfico más adecuado (en este caso el de columnas). El resultado será:

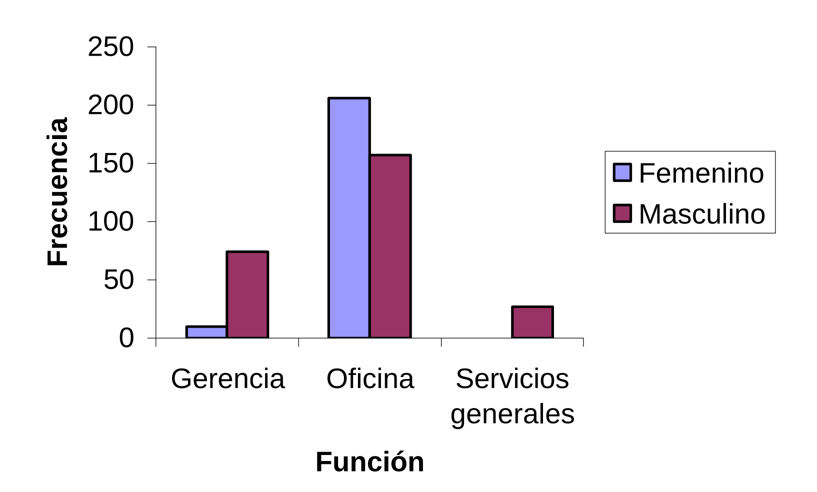


Figura 15 Gráfico de columnas: Función por Sexo

1.2 Procedimientos para variables cuantitativas

Para variables cuantitativas hay una mayor variedad de procedimientos disponibles. Es necesario distinguir los procedimientos relativos a las variables discretas y contínuas, tanto para los casos en que se desee hacer un análisis de una variable cuantitativa en función de una cualitativa (por ejemplo, Salario Anual por Sexo). En los siguientes itens se presentan los principales procedimientos.

1.2.1 Procedimientos para variables cuantitativas discretas

Se la variable a analizar es discreta, por ejemplo Años de Educación (AñosEd) en la planilla Empleados, el procedimiento puede ser semejante al utilizado para la variable SEXO (ítem 1.1.1), sin embargo al construir la tabla dinámica, Excel irá a seleccionar "Suma de

AñosEd) como acción (porque los valores de la variable son números), y se necesitará modificar eso para conteo de los valores: CONTAR de AñosEd, siguiendo el procedimiento que se ilustró en la Figura 8. Luego de realizar los ajustes necesarios, se obtendrá la siguiente tabla dinámica:

Contar de AñosEd	
AñosEd	Total
8	53
12	190
14	6
15	116
16	59
17	11
18	9
19	27
20	2
21	1
Total general	474

Figura 16 Tabla de frecuencia: Años de educación

A partir de la información mostrada en la Figura 16, se puede generar el siguiente gráfico:

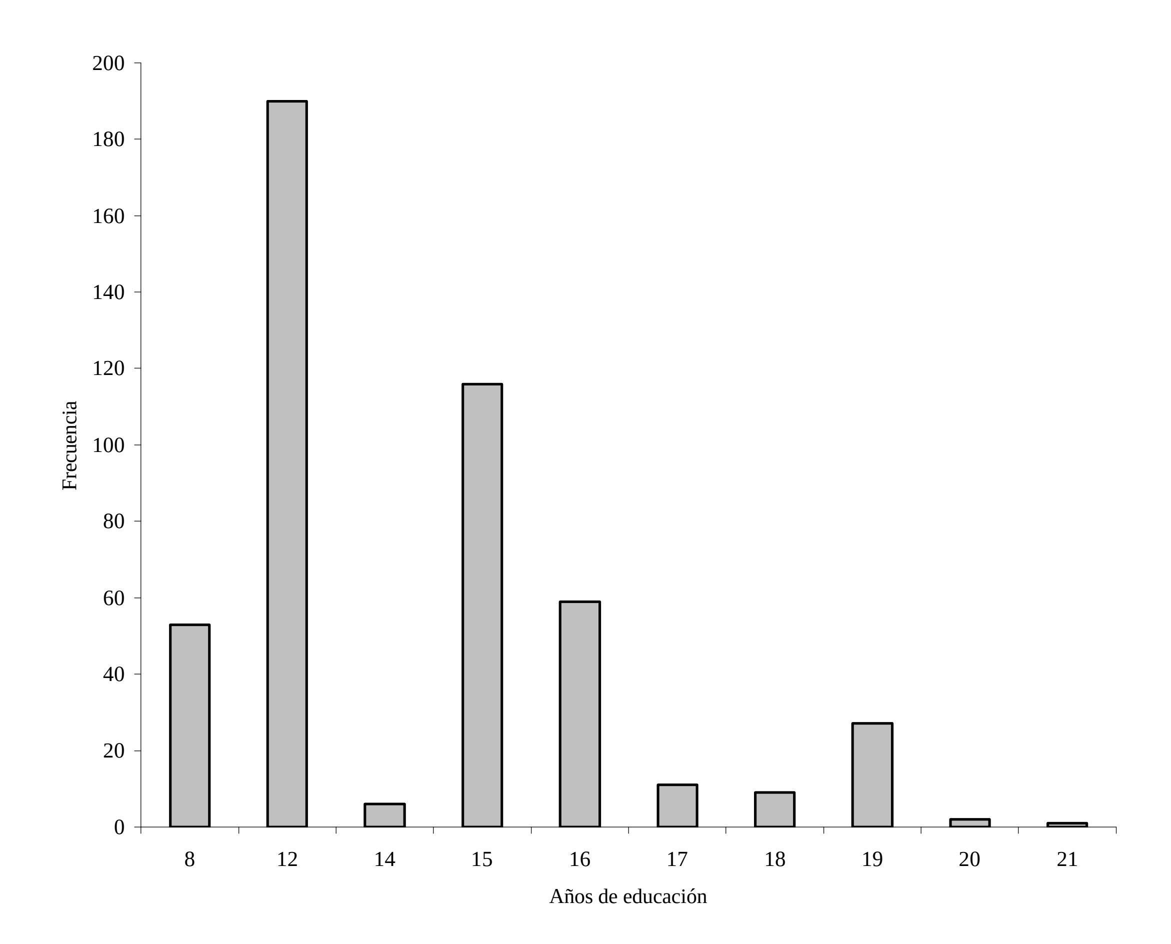


Figura 17 Gráfico de barras para la variable Años de educación

1.2.2 Procedimientos para variables cuantitativas contínuas

Si la variable de interés es de naturaleza cuantitativa contínua, como Salario anual "SalarioA" en la planilla Empleados, se puede tener interés en construir una tabla de frecuencias agrupada en clases. Por esta razón NO SE PUEDE UTILIZAR la tabla dinámica de Excel: como los valores de una variable contínua se repiten poco (o no se repiten), se tendría una tabla inmensa (probablemente con centenares de líneas en el caso de la planilla Empleados). A continuación se explicará cómo construir la tabla utilizando algunas funciones existentes en Excel, como MÁXIMO, MINIMO, CONTAR.SI, entre otras.

1.2.2.1 Tabla de frecuencias agrupada en clases

Los pasos a seguir son similares a los presentados en la salón de clase.

1. Determinar el rango o amplitud del conjunto de datos

Para obtener el rango, se tienen que identificar los extremos del conjunto de datos, o sea, sus valores máximo y mínimo. Se iniciará por el mínimo. Seleccione una celda donde desea que el resultado sea colocado: por ejemplo la celda L2. Seleccione esta celda con el cursor. Observe que en la barra de herramientas de Excel hay un botón llamado PEGAR FUNCION *fx* , vea la Figura 18.

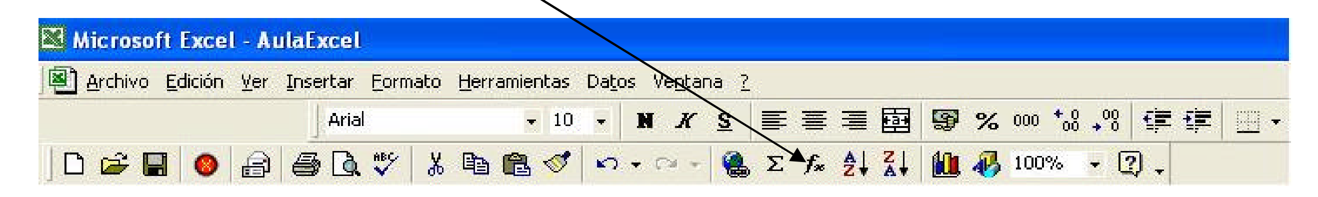


Figura 18 Barra de herramientas de Excel : "Pegar Función"

Si presiona *fx* surgirá la pantalla vista en la Figura 19.

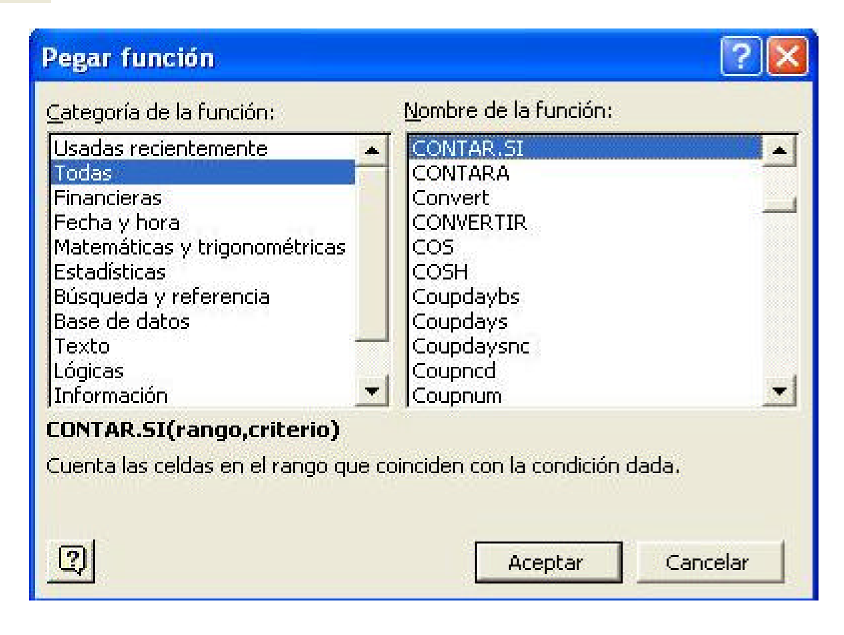


Figura 19 Funciones disponibles en Excel

A la izquierda (en la Figura 19) se puede ver la categoría de la función, y a la derecha los nombres de las funciones disponibles, con una breve descripción de cada una de ellas. Una de las categorías se llama: “Usadas recientemente”, que registra las últimas funciones aplicadas por el usuario, en cualquier planilla. Estamos especialmente interesados en las categorías “Matemáticas y trigonométricas”, “Lógicas” y, obviamente “Estadísticas”. Las funciones MAX y MIN se encuentran en esta última categoría. Seleccionando “Estadísticas” vamos a obtener el resultado de la Figura 20.

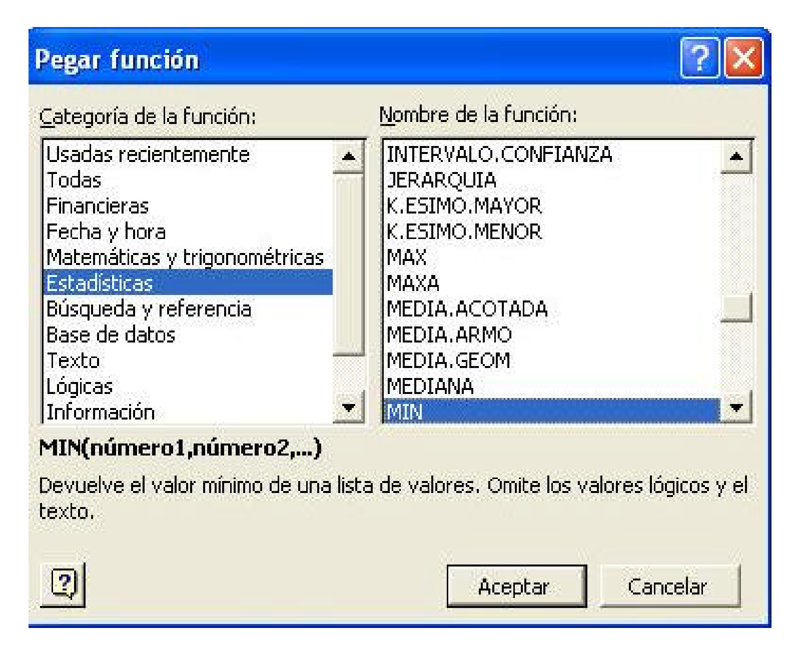


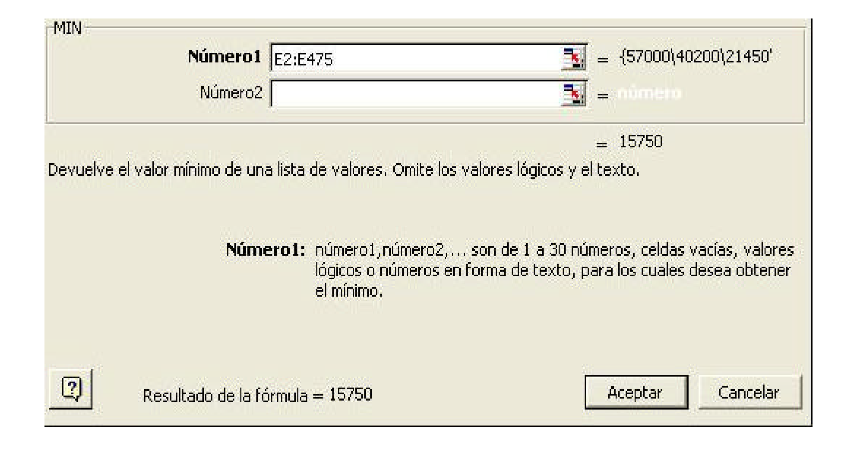
Figura 20 Funciones estadísticas

Luego de seleccionar “Estadística”, basta buscar la función MIN: y observe la descripción en la parte inferior. También se puede pedir ayuda a Excel sobre la descripción detallada de las funciones.

Buscando detenidamente, se encontrarán otras funciones estadísticas muy útiles en el análisis de una variable cuantitativa contínua, tal es el caso de:

PROMEDIO (media aritmética),
MEDIANA (mediana)
VAR (varianza) y DESVEST (desviación estándar),
PERCENTIL, etc.

Una vez seleccionada la función, basta presionar ACEPTAR y para el caso de MIN se tendrá:



Se necesita seleccionar el intervalo de valores de la variable, para que Excel pueda identificar el valor mínimo. Observe el resultado de la fórmula, y también el botón que indica la ayuda (que podrá necesitarse para otras funciones más complejas)

Figura 21 Función MIN

El mismo resultado podría ser obtenido simplemente digitando la fórmula directamente en la celda: =MIN(E2:E475) . Pueden ser utilizadas mayúsculas o minúsculas.

Para encontrar el valor máximo se puede realizar un proceso análogo utilizando la función MAX, pero colocando el resultado en otra celda, L3 por ejemplo. Para calcular el intervalo se puede colar una fórmula en la celda L4, haciendo la sustracción entre máximo y mínimo. Los resultados pueden observarse a continuación:

Salario Anual	
Mínimo	15750
Máximo	135000
Intervalo	119250

Figura 22 Intervalo del salario anual

El menor salario anual es Q 15,750.00 y el mayor de Q 135,000.00, resultando en una amplitud de Q 119,250. Esta amplitud es la se necesita para la construcción de la distribución en clases del conjunto de datos.

2. Dividir el rango en un número conveniente de clases

Usualmente se define el número de clases (NC), utilizando la ecuación de Sturges, NC = 1+3.33 * $\log_{10}$ (n), como en nuestro caso n = 474 personas, tenemos que el número de clases es aproximadamente igual a 10. De acuerdo a esto, la amplitud (o ancho de clase) sería igual a 11,925.

3. Establecer los límites de las clases

Podemos definir valores diferentes para la amplitud de las clases y el valor inicial, para este último, debe considerarse que sea menor de 15,750 (mínimo) y la amplitud garantice que el valor máximo sea incluido en el conjunto de datos. Seleccionando una amplitud de 12,000 y un valor inicial de 15,000, teniendo en mente que la tabla tendrá 10 clases, el resultado será (los límites también pueden ser calculados a través de Excel):

[15000	27000)
[27000	39000)
[39000	51000)
[51000	63000)
[63000	75000)
[75000	87000)
[87000	99000)
[99000	111000)
[111000	123000)
[123000	135000)

Observe que los valores del conjunto (del mínimo al máximo) formarán parte de las clases, la ya que la última clase el límite superior también fue incluido. Podemos colocar los límites de cada clase en celdas de Excel, para posteriormente construir un gráfico.

4. Determinar las frecuencias de cada clase.

Este es el paso más difícil, pero podemos resolverlo utilizando la función CONTAR.SI. Esta función cuenta cuántos valores en un determinado intervalo de datos atienden a un criterio establecido. El establecimiento del criterio no permite sin embargo, que sean incluidos dos límites, por ejemplo, contar todos los valores que son mayores que 15,000 y menores que 27,000. Además de eso, no es posible utilizar otras celdas de la planilla al definir el criterio. Para el caso en mención, necesitaremos insertar la función en un celda al lado de aquella

donde se encuentran los límites de la primera clase. Si presiona surgirá la pantalla vista en la Figura 20, busque por ESTADÍSTICAS, resultando la figura 22.

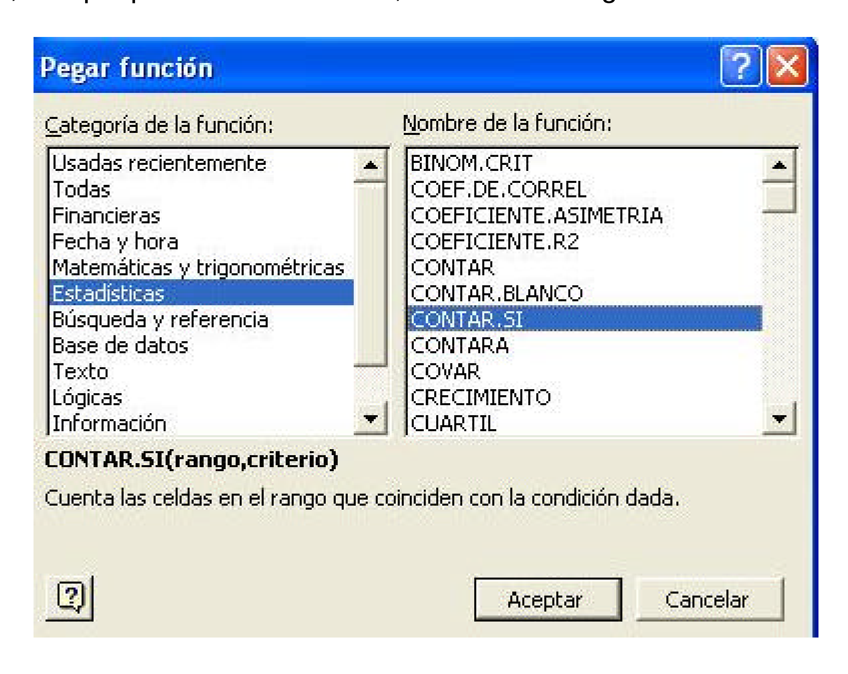
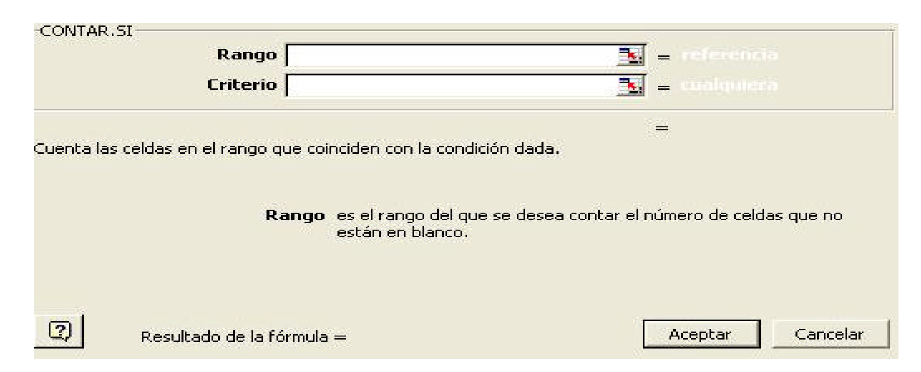


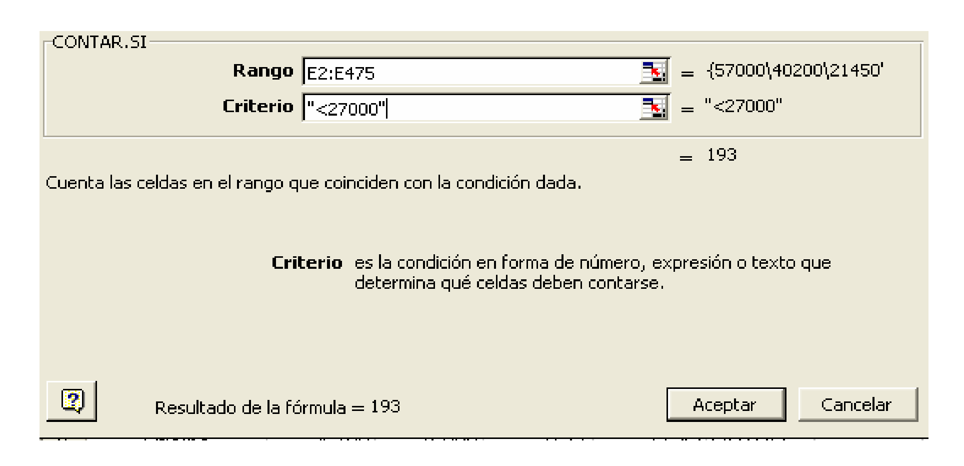
Figura 22 Función CONTAR.SI

Presionando ACEPTAR causará el aparecimiento de la Figura 23



Es necesario definir el rango de datos: los datos de interés están en las celdas E2 a E475 en la planilla Empleados. Y el criterio será <27000, o sea, el límite superior de la primera clase. Vea la Figura 24

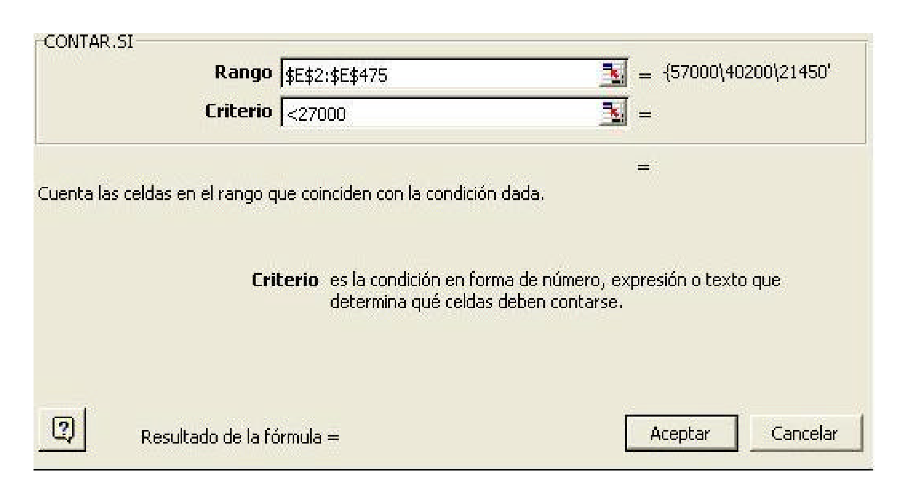
Figura 23 Rango de datos y criterio para función CONTAR.SI – parte 1



Si presiona ACEPTAR, Excel mostrará que hay 193 salarios inferiores a 27,000. Por otra parte, si quisiera arrastrar esta celda para abajo, para aplicarla a las otras celdas, tendría problemas

Figura 24 Rango de datos y criterio para la función CONTAR.SI – parte 2

Como el rango E2:E475 no es una referencia absoluta, al arrastrar la fórmula hacia abajo, el rango de los datos también se irá a mover: E3:ÑE476, E4:E477, y así sucesivamente, lo que los puede llevar a valores incorrectos. Para evitar eso, basta con marcar el rango en la Figura 24, y presionar la tecla F4 en el teclado del computar que estuviere usando: el intervalo será considerado como referencia absoluta, y al arrastrar las fórmulas, sus celdas permanecerán las mismas. Vea el efecto en la Figura 25.



Este procedimiento (utilización de la tecla F4) puede ser realizado en cualquier situación en que sea necesario establecer una referencia absoluta. Presionando ACEPTAR, y arrastrando la fórmula hacia abajo:

Figura 25 Rango de datos y criterio para función CONTAR.SI – parte 3

<table>
<tr><th>Clases</th><th>Frecuencia</th></tr>
<tr><td>15000 | --27000</td><td>193</td></tr>
<tr><td>27000 | --39000</td><td>193</td></tr>
<tr><td>39000 | --51000</td><td>193</td></tr>
<tr><td>51000 | --63000</td><td>193</td></tr>
<tr><td>63000 | --75000</td><td>193</td></tr>
<tr><td>75000 | --87000</td><td>193</td></tr>
<tr><td>87000 | --99000</td><td>193</td></tr>
<tr><td>99000 | --111000</td><td>193</td></tr>
<tr><td>111000 | --123000</td><td>193</td></tr>
<tr><td>123000 | --135000</td><td>193</td></tr>
</table>

Figura 26 Resultados de CONTAR.SI

Observe que el mismo valor fue repetido para todas las clases. Eso ocurrió porque el criterio (“<27000) permaneció constante para todas. Se necesita cambiar manualmente los criterios para cada clase. Por ejemplo, en la celda al lado de la clase 27000 – 39000, el 27000 (que fue arrastrado de la celda anterior) debe ser sustituido por 39000. En la celda siguiente, debe ser sustituido por 51000, y así sucesivamente. Para hacer eso, basta colocar el cursor sobre la celda deseada y cambiar el valor del criterio, que está entre comillas. En la celda referente a la última clase, es necesario adicionar el símbolo = después el < del criterio (además de cambiar el valor para 135000), para garantizar la incorporación del valor máximo.



Figura 27 Barra de fórmulas de Excel –CONTAR.SI: sustitución del criterio

Luego de realizar este procedimiento tendrá:

Clases	Frecuencia
15000 | --27000	193
27000 | --39000	366
39000 | --51000	404
51000 | --63000	435
63000 | --75000	455
75000 | --87000	465
87000 | --99000	469
99000 | --111000	473
111000 | --123000	473
123000 |--|135000	474

La tabla mostrada al lado registró las frecuencias ACUMULADAS hasta las respectivas clases. Esa información puede ser importante, pero estamos interesados en obtener las frecuencias individuales de cada clase. Nuevamente necesitamos modificar las fórmulas, sustrayendo de cada una, a partir de la segunda clase, las frecuencias de todas las clases anteriores. Por ejemplo para la clase 27000 | --39000, es necesario sustraer de su frecuencia (<366), las frecuencia anterior (193). Vea la expresión en la Figura 29.

Figura 28 Resultado de CONTAR.SI –1ª modificación

Realizando las modificaciones, el resultado será:

Clases	Frecuencia
15000 | --27000	193
27000 | --39000	173
39000 | --51000	38
51000 | --63000	31
63000 | --75000	20
75000 | --87000	10
87000 | --99000	4
99000 | --111000	4
111000 | --123000	0
123000 | --135000	1

Ahora si, las frecuencias están correctas. Podemos observar, como era esperado, una mayor frecuencia de salarios anuales más bajos, hasta Q 39,000. Esta tabla puede ser usada para construir un histograma. Seleccionando la tabla, escogiendo el gráfico de columnas, y reduciendo el espacio entre barras a cero, entre otros ajustes, se obtendrá un histograma tal como se muestra en la Figura 30.

Figura 29 Resultado de CONTAR.SI –2ª modificación

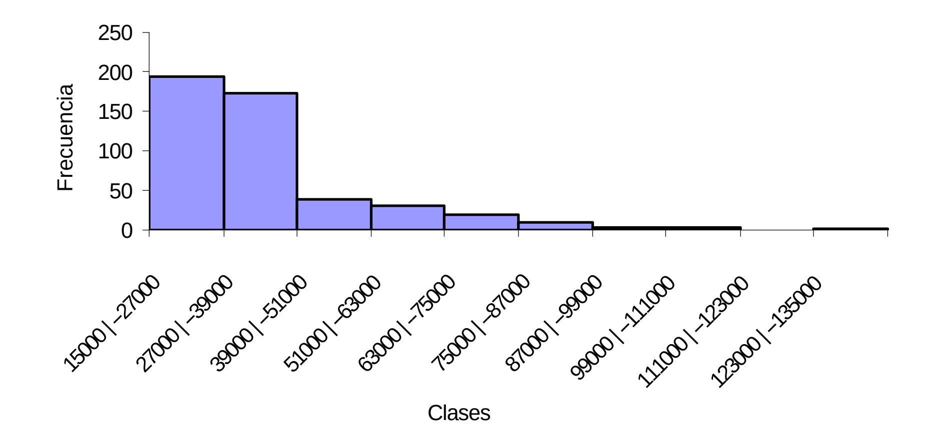


Figura 30 Histograma de los salarios anuales (agrupados en clases)

1.3 Uso de las funciones de MS Excel para obtener medidas de tendencia central, de dispersión, de oblicuidad y curtosis.

En esta sección se mostrará como las funciones[3] de Excel pueden ser utilizadas para calcular cada una de las medidas de tendencia central, dispersión, oblicuidad y curtosis**.** Además se discutirán dos medidas de localización no centrales (percentiles y cuartiles).

1.3.1 Funciones de Excel para medidas de tendencia central

Entre las funciones que se encuentran disponibles para obtener medidas de tendencia central están:

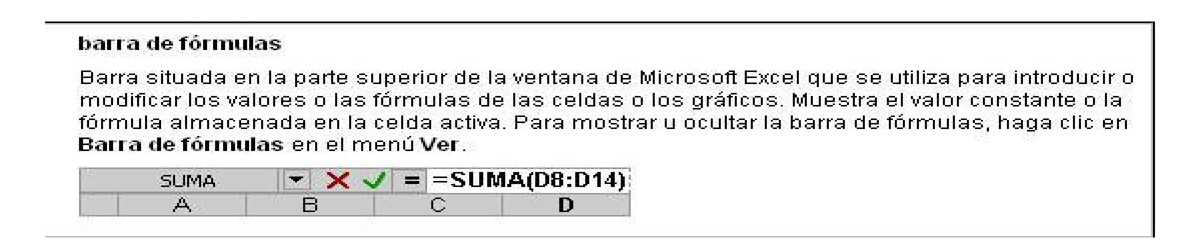
Función	Descripción
MEDIA.ARMO	Calcula la media armónica de un conjunto de números positivos: el recíproco de la media aritmética de los recíprocos.
MEDIA.GEOM	Calcula la media geométrica de una matriz o de un rango de datos positivos.
MEDIANA	Calcula la mediana de los números. La mediana es el número que se encuentra en medio de un conjunto de números, es decir, la mitad de los números es mayor que la mediana y la otra mitad es menor.
MODA	Calcula el valor que se repite con más frecuencia en una matriz o rango de datos.
PROMEDIO	Devuelve el promedio (media aritmética) de los argumentos.

Para introducir una fórmula que contenga una función, siga los siguientes pasos:

1. Haga clic en la celda en que desee introducir la fórmula.

[3] Las funciones son fórmulas predefinidas que ejecutan cálculos utilizando valores específicos, denominados argumentos, en un orden determinado o estructura.

2. Para iniciar la fórmula con la función, haga clic en **Modificar fórmula** = en la barra de fórmulas.


barra de fórmulas

Barra situada en la parte superior de la ventana de Microsoft Excel que se utiliza para introducir o modificar los valores o las fórmulas de las celdas o los gráficos. Muestra el valor constante o la fórmula almacenada en la celda activa. Para mostrar u ocultar la barra de fórmulas, haga clic en **Barra de fórmulas** en el menú **Ver**.

Figura 31 Barra de fórmulas

3. Haga clic en la flecha hacia abajo que aparece junto al cuadro **Funciones** (el botón que aparece al lado derecho de la función SUMA en la Figura 31)

4. Haga clic en la función que desee agregar a la fórmula. Si la función no aparece en la lista, haga clic en **Más funciones** para obtener una lista de las funciones adicionales. También puede seleccionar **Pegar función** *fx* y seleccionar la función requerida.

5. Introduzca los argumentos (valores que utiliza una función para ejecutar las operaciones o cálculos)

6. Una vez completa la fórmula, presione ACEPTAR.

1.3.2 Funciones de Excel para medidas de dispersión

Entre las funciones que se encuentran disponibles para obtener medidas de dispersión están:

Función	Descripción
DESVEST	Calcula la desviación estándar de una muestra. DESVEST parte de la hipótesis de que los argumentos representan la muestra de una población. La desviación estándar se calcula utilizando el método "insesgado" o "n-1".
DESVESTP	Calcula la desviación estándar de la población total determinada por los argumentos. DESVESTP parte de la hipótesis de que los argumentos representan la población total.
DESVIA2	Devuelve la suma de los cuadrados de las desviaciones de los puntos de datos a partir de la media de la muestra.
VAR	Calcula la varianza de una muestra. La función VAR parte de la hipótesis de que los argumentos representan una muestra de la población. Si sus datos representan la población total, utilice **VARP**

	para calcular la varianza.

1.3.3 Funciones de Excel para medidas de oblicuidad (o de asimetría)

La función **COEFICIENTE.ASIMETRÍA** calcula la asimetría de una distribución. Esta función caracteriza el grado de asimetría de una distribución con respecto a su media.

La asimetría positiva indica una distribución unilateral que se extiende hacia valores más positivos. La asimetría negativa indica una distribución unilateral que se extiende hacia valores más negativos.

1.3.4 Funciones de Excel para medidas de curtosis

La función **CURTOSIS** calcula el coeficiente de curtosis de un conjunto de datos. La curtosis representa la elevación o achatamiento de una distribución, comparada con la distribución normal. Una curtosis positiva indica una distribución relativamente elevada (leptocurtica), mientras que una curtosis negativa indica una distribución relativamente plana (platicurtica). Valores cercanos a cero, indican una distribución mesocurtica.

1.3.5 Cálculo de percentiles y cuartiles

El percentil de orden 100p (P_{100p}) de un conjunto de valores dispuestos en orden creciente, es un valor tal que (100p)% de las observaciones son menores o iguales a él, y 100 (1-p)% son mayores o iguales a él ($0 < p < 1$).

4 El percentil de orden 50 (P_{50}) es igual a la mediana.
4 Los percentiles de orden 25, 50 y 75 representados por Q_1, Q_2 y Q_3 son llamados cuartiles.

CÁLCULO DEL PERCENTIL DE ORDEN 100 (P_{100p}) PARA DATOS NO AGRUPADOS

np entero:

$$P_{(100p)} = \frac{x_{[np]} + x_{[np+1]}}{2}$$

np no entero:

$$P_{(100p)} = x_{[\text{int}(np)+1]}$$

siendo int (.) la función que aproxima un número para abajo hasta el entero más próximo. Por ejemplo: int (1.9) = 1, int (1.5) =1, int (1.2) = 1.

Ejemplo

En la siguiente tabla, se presentan los valores correspondientes a la producción (en gramos) de hule seco por sangría, por planta de hule, en el área A de la hacienda "Río Blanco".

10.2	10.2	10.3	10.6	10.8	11.0	11.6	11.8	11.9	12.0
12.1	12.6	12.6	12.8	12.8	13.0	13.1	13.2	13.4	13.5
14.0	14.9	15.2	15.3	15.3	15.4	15.8	16.0	16.2	16.3
16.9	17.7	18.1	18.3	18.4	18.7	19.6	19.8	19.9	20.0

20.3	20.3	21.9	22.0	22.2	22.4	22.8	23.2	23.5	23.8
24.2	24.5	24.6	24.9	25.1	25.5	26.0	26.3	26.8	28.1

Los datos están localizados en la planilla Hule del archivo AulaExcel1.xls, disponible en: C:\\Mis Documentos\Estadística\

Con esta información establecer:

a) El valor de la producción que separa el 10% de las plantas de hule más productivas.

n = 60 datos
p = 0.90 (percentil 90)
np = 54

$$P_{90} = \frac{x_{[54]} + x_{[55]}}{2} \qquad P_{90} = \frac{24.9 + 25.1}{2} = 25\,\text{gr}$$

Para encontrar la solución a este ejercicio, utilizando Excel, se utiliza **Pegar función** f_x y se selecciona la categoría ESTADÍSTICAS, como se ilustra en la Figura 32:

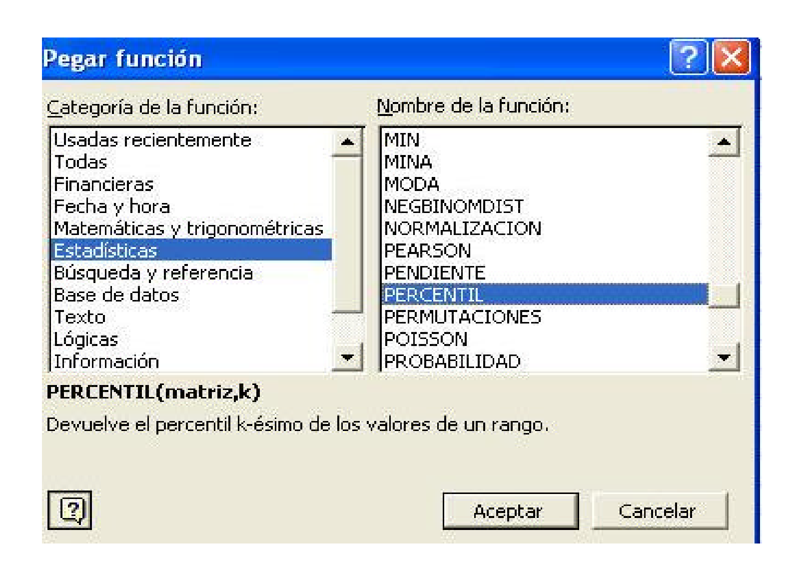


Figura 32 Función percentil

Luego de presionar ACEPTAR, se desplegará la siguiente pantalla:

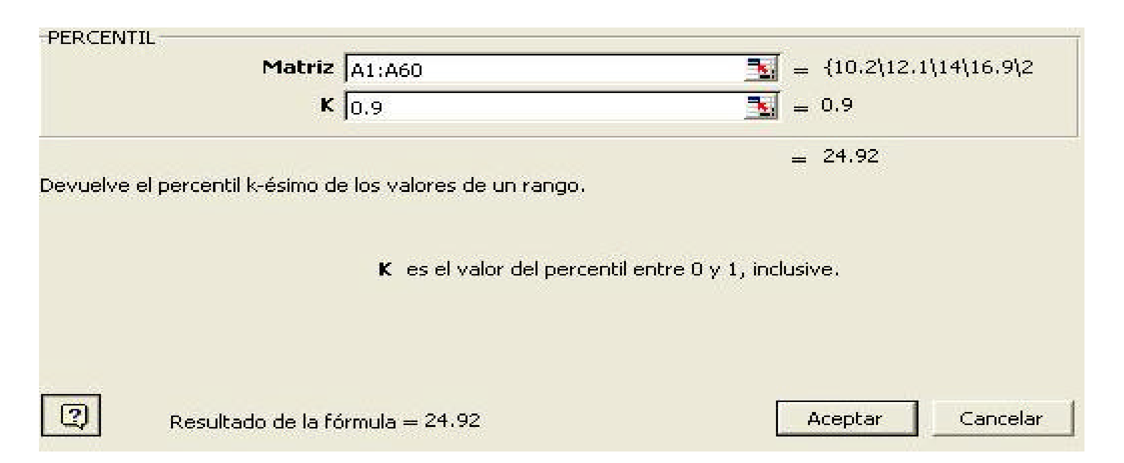


Figura 33 Resultados del cálculo de percentiles

Observe en la Figura 33 que en el primer espacio en blanco, aparece la palabra MATRIZ, allí se debe indicar dónde se encuentra el conjunto de datos, recuerde que deben colocarse en una sola columna (por cada variable), luego en el segundo espacio aparece la letra **K**, donde se debe indicar el valor del percentil (debe estar entre 0 y 1). Para el caso en estudio, nos interesa el percentil 90 (p=0.9). El resultado es: 24.92

b) Calcular el percentil de orden 2.5

n = 60 datos
p = 0.025
np = 1.5

$$P_{2.5} = x_{[\text{int}(1.5)+1]} = x_{(2)} \quad P_{2.5} = 10.2 \text{ gr}$$

c) Calcular el percentil de orden 97.5

n = 60 datos
p = 0.975
np = 58.5

$$P_{97.5} = x_{[\text{int}(58.5)+1]} = x_{(59)} \qquad P_{97.5} = 26.8 \text{ gr}$$

- **Cuartiles y desviación cuartil**

Cuando se calcula la mediana de una serie de datos cuantitativos, éstos se ordenan y la mediana los divide en dos grupos con la misma cantidad de elementos: hay un grupo inferior y otro superior. Cada uno de esos grupos, que ya están ordenados, tienen a su vez una mediana. La mediana del grupo inferior se llama *primer cuartil*, denotado como Q_1 y la mediana del grupo superior se llama tercer *cuartil*, denotado como Q_3.

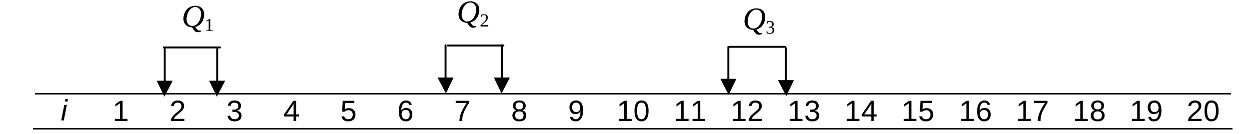
El *segundo cuartil* (Q_2)es la mediana original de la serie completa de datos. Véase que la función de los cuartiles es dividir los datos originales en cuatro grupos con la misma cantidad de datos cada uno. Así, habrá un primer grupo que contiene al 25% de los datos y que va desde el menor de los datos hasta Q_1. El segundo grupo contiene al 25% de los datos y va de Q_1 a la mediana. El tercer grupo contiene al 25% de los datos y va de la mediana a Q_3. Finalmente, el cuarto grupo contiene también al 25% de los datos y va de Q_3 hasta el mayor de los datos.

La *desviación cuartílica* se define como:

$$Q = \frac{Q_3 - Q_1}{2}$$

Ejemplo

Considerando los datos del siguiente ejemplo, referente a las notas obtenidas por un grupo de 20 estudiantes universitarios (los datos está disponibles en la planilla Estudiantes del archivo AulaExcel1.

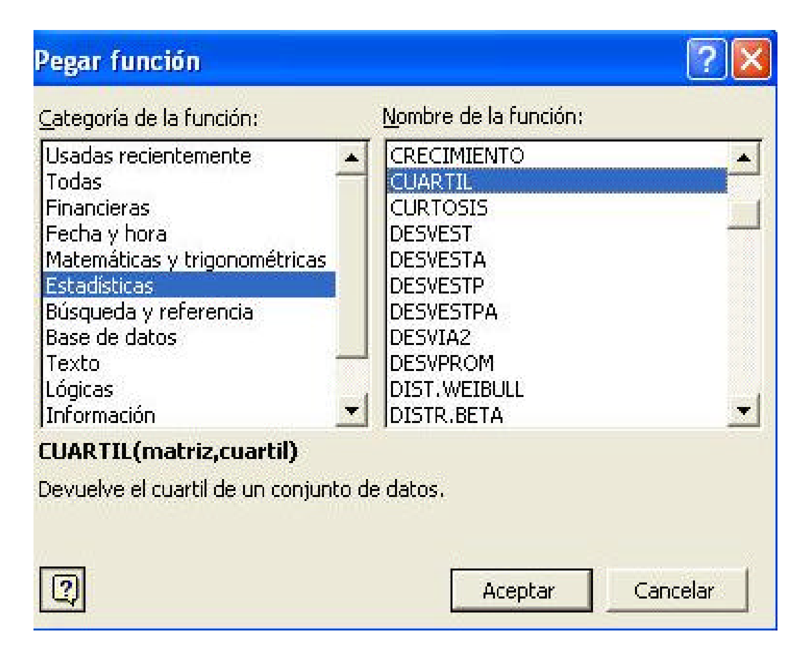


$x_{(i)}$	15	45	47	53	58	58	60	62	67	74	75	78	80	80	81	85	85	85	90	92

Los cuartiles son: Q_1 = 58 puntos, Md = 74.5 puntos, Q_3 = 83 puntos. Entonces la desviación cuartílica es:

$$Q = \frac{83 - 58}{2} = 12.5 \text{ puntos}$$

Puede comprobar estos resultados, utilizando **Pegar función** , como se ilustra en la Figura 34.



Luego de seleccionar la función CUARTIL, dé un clic en el botón ACEPTAR, y se desplegará una pantalla similar a la mostrada en la Figura 35

Figura 34 Función Cuartil

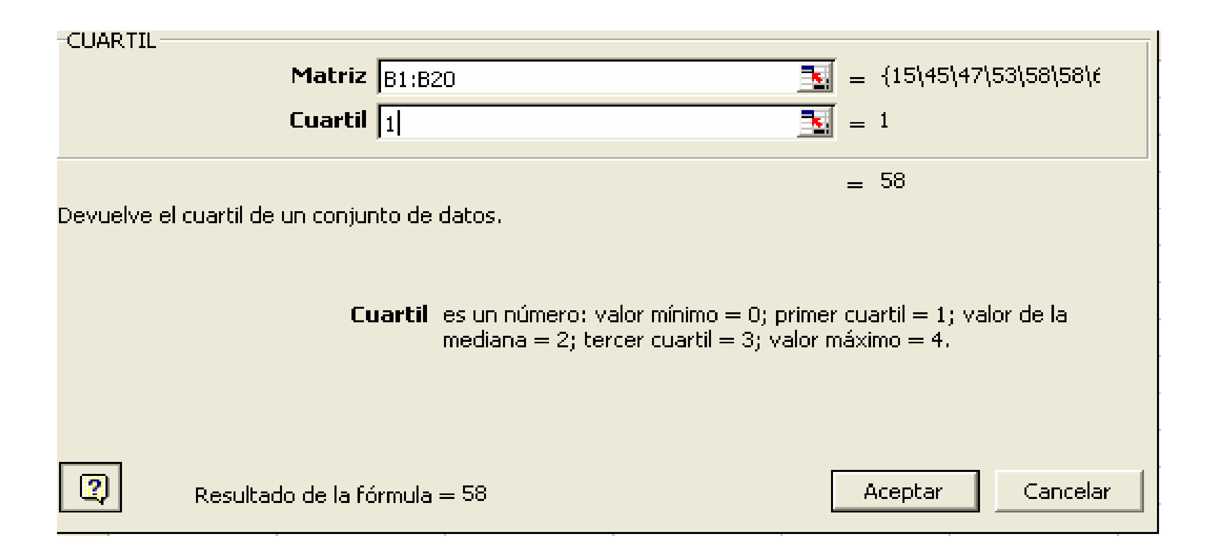


Figura 35 Resultado del cálculo del primer cuartil

Observe en la Figura 35 que en el espacio referente a Cuartil, debe indicar qué cuartil está interesado en obtener, si coloca 1, obtendrá el primer cuartil, si es 2, el segundo cuartil y si es 3, el programa calculará el tercer cuartil.

1.4 Uso de la herramienta ANÁLISIS DE DATOS para obtener estadísticas descriptivas

Además de (o en vez de) utilizar diversas funciones para calcular estadísticas descriptivas, se puede utilizar la herramienta ANÁLISIS DE DATOS para obtener simultáneamente un conjunto de estadísticas descriptivas. La ventaja de esta herramienta, reside en que, es más fácil de utilizar, debido a que se pueden obtener varias estadísticas a través de apenas una operación. Una desventaja es que, si los datos son modificados luego de haber utilizado la herramienta, toda la secuencia de etapas tendrá que ser repetida con los nuevos datos.

Para poder utilizar esta herramienta debe previamente revisar si su computadora cuenta con ella, haciendo los siguiente:

- Ingrese al menú HERRAMIENTAS y al desplegarlo, verifique que se encuentre la opción ANÁLISIS DE DATOS, si no se encuentra, seleccione la opción COMPLEMENTOS dentro del menú HERRAMIENTAS y seleccione: HERRAMIENTAS PARA ANÁLISIS Y HERRAMIENTAS PARA ANÁLISIS VBA, tal como se ilustra en la Figura 36

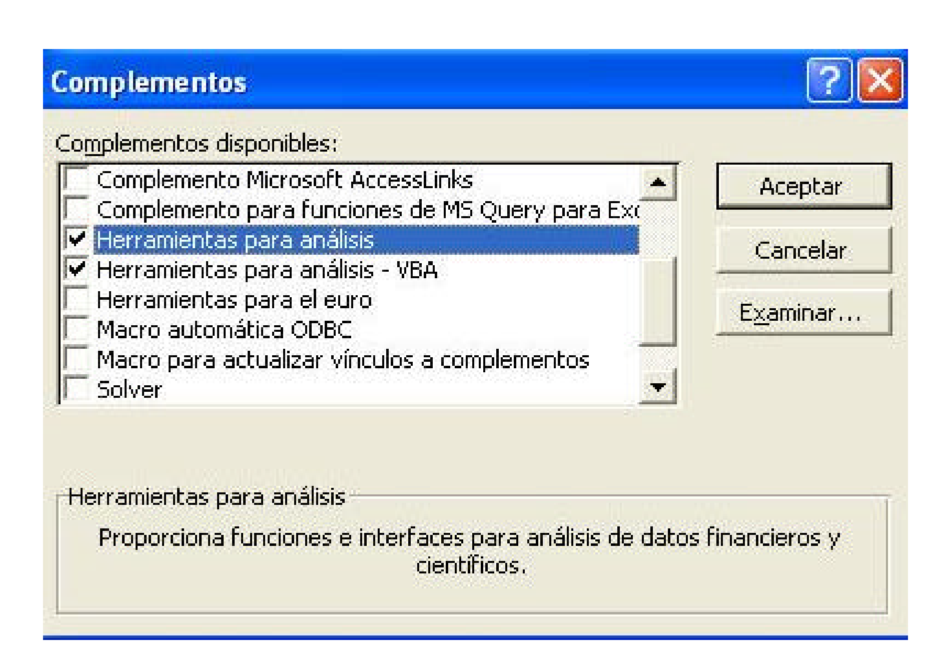


Figura 36
Selección de las Herramientas para análisis estadístico de datos en Excel.

Luego de un clic en ACEPTAR y verifique que en el menú HERRAMIENTAS aparezca la opción ANÁLISIS DE DATOS, la cual es de gran utilidad para diversos análisis estadísticos (vea la figura 37)

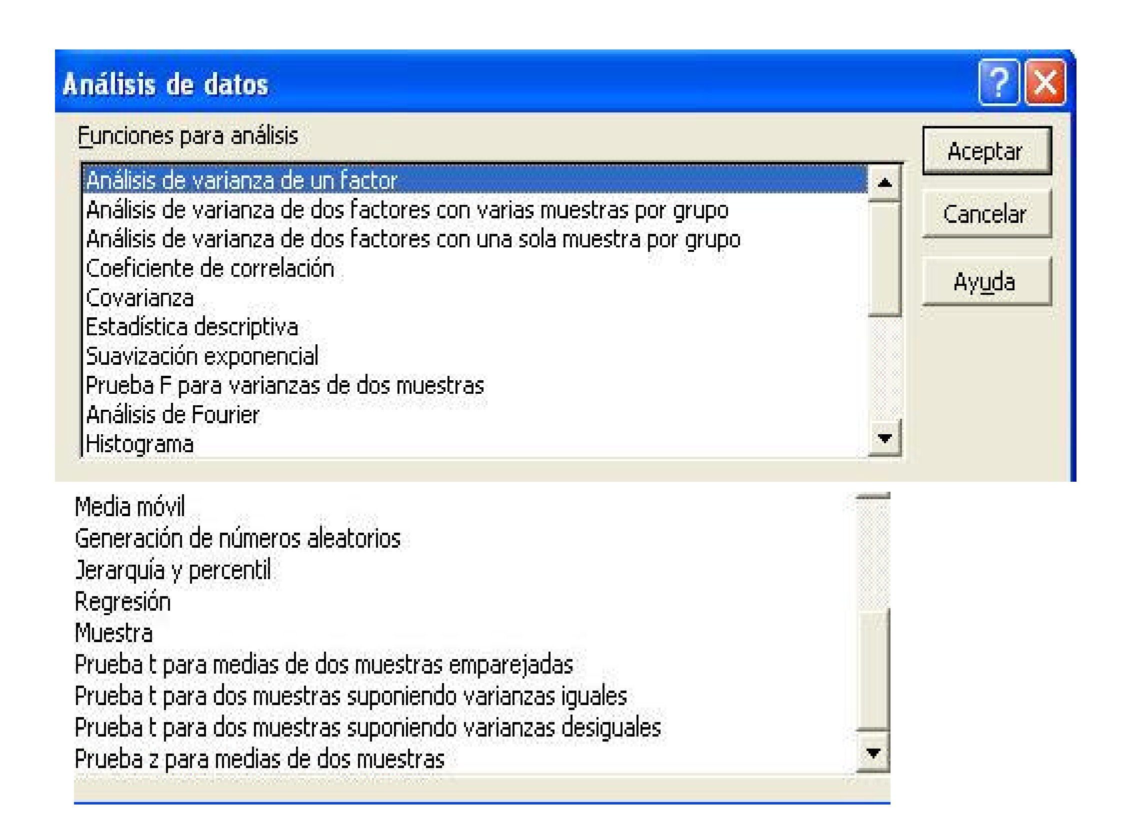


Figura 37 Funciones para análisis estadístico disponibles en Excel.

A continuación se mostrará un ejemplo para el uso de la herramienta ANÁLISIS DE DATOS.

En un levantamiento forestal fue medido el diámetro (expresado en centímetros) de algunos árboles de dos florestas, obteniéndose los siguientes valores (disponibles en la hoja Florestas del archivo AulaExcel1):

FLORESTA A

16	50	13	8	5	77	93	27	57	28	24	16
49	60	7	5	9	30	8	51	41	33	62	35
9	49	31	107	27	56	26	55	10	18	7	24
17	63	11	34	19	12	40	28	6	19	10	50
16	29	22	10	17	36	42	134	7	10	29	14
12	12	29	76	10	106	52	43	17	16	51	19
21	96	87	29	77	6	9	21	18	6	15	161
32	12	16	29	7	20	37	76	47	6	17	35
30	44	13	56	112	38	15	56	17	34	43	6
52	42	35	25	31	127	9	21	5	154	13	7

FLORESTA B

38	43	32	18	47	33	38	27	50	34	34	31
28	31	46	27	33	33	38	24	33	23	16	42
22	26	27	32	23	46	30	9	36	47	21	61
34	37	36	30	41	16	7	33	50	11	27	7
23	27	38	23	25	33	30	36	27	32	23	25
30	23	40	15	23	47	35	39	41	46	35	30
42	23	43	35	28	31	35	33	30	30	49	34
48	29	29	30	21	32	28	31	36	22	26	41
33	25	42	30	22	17	40	39	36	55	29	40
42	29	37	29	32	49	17	22	48	31	42	38

Con esta información se le solicita:

a) Calcular las medidas de tendencia, dispersión, asimetría y curtosis para cada conjunto de datos.

b) Construir un histograma y un polígono de frecuencias para cada conjunto de datos.

c) ¿Existen observaciones extremas en esos conjuntos de datos? (construya un *box plot* para los datos de cada floresta)

d) Describa la forma de la distribución en las dos florestas.

Solución:

Es necesario recordar, que los datos referentes a cada variable (en este caso, a cada floresta) deben estar ubicados en una sola columna o una sola fila. Inicialmente se explicará cómo obtener las estadísticas descriptivas para la Floresta A, siguiendo los siguientes pasos:

a) En el menú HERRAMIENTAS, seleccione la opción ANÁLISIS DE DATOS

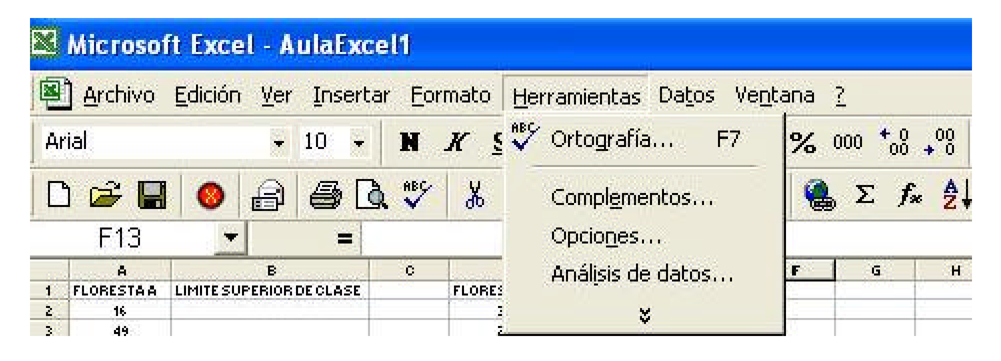


Figura 38. Herramienta para análisis de datos

b) Luego se desplegará el siguiente la herramienta para ANÁLISIS DE DATOS, donde se deberá seleccionar la opción ESTADÍSTICA DESCRIPTIVA, tal como se ilustra en la Figura 39

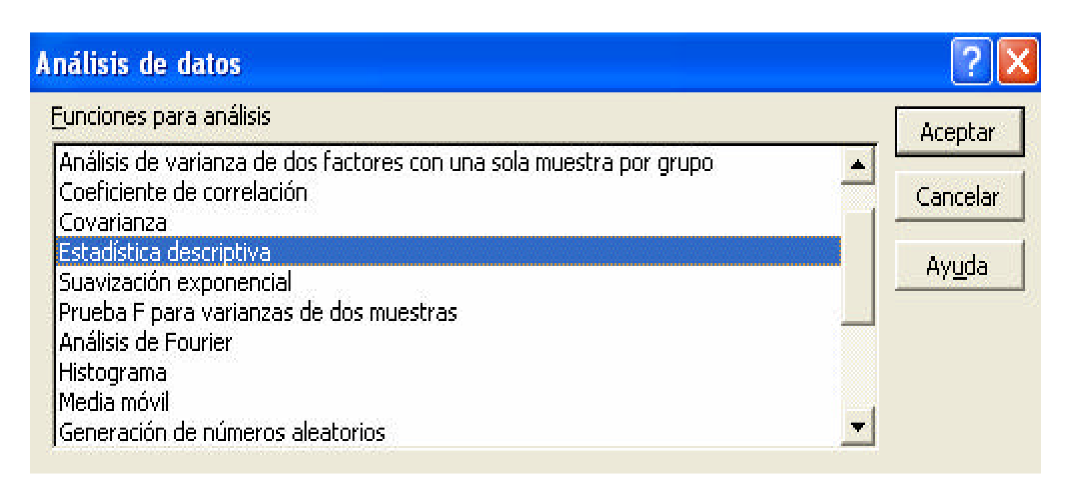


Figura 39 Función Estadística descriptiva

c) Luego de seleccionar la opción ESTADÍSTICA DESCRIPTIVA y presionar ACEPTAR se desplegará la siguiente pantalla:

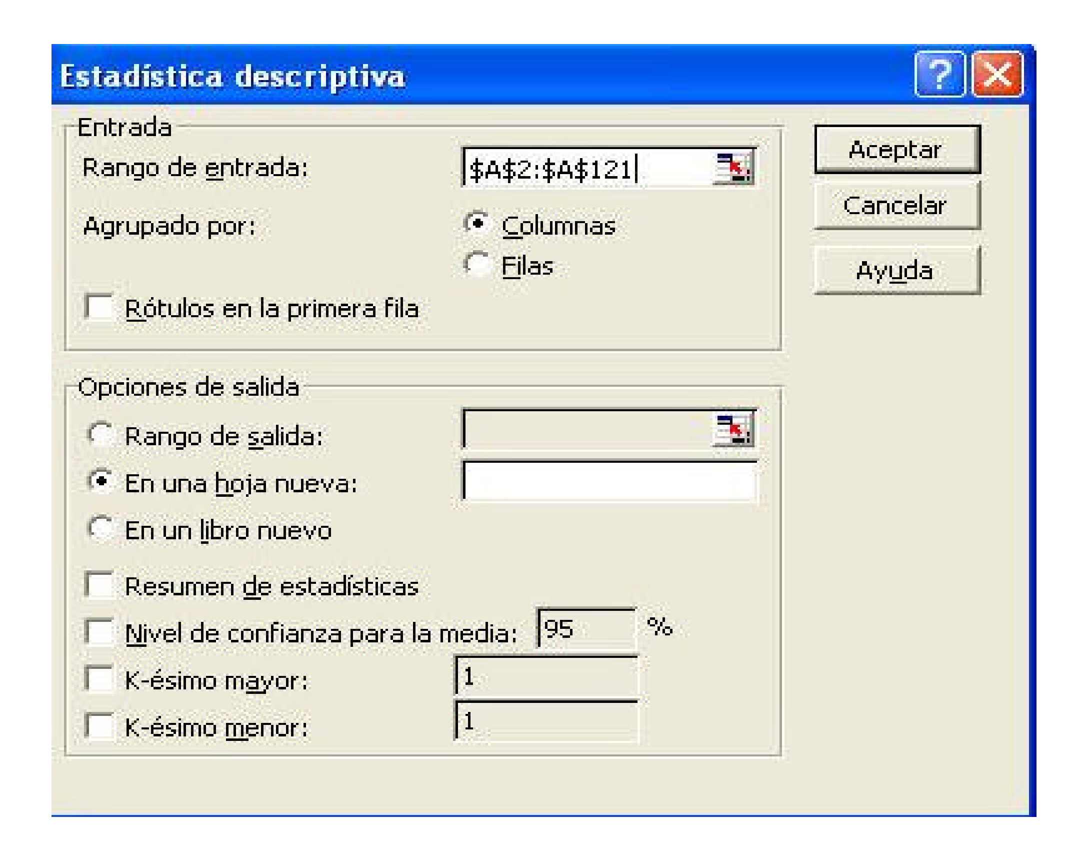


Figura 40 Opciones de la función ESTADÍSTICA DESCRIPTIVA

En esta pantalla se debe marcar el rango de entrada de los datos e indicar si están colocados en una columna o en una fila, si decidió colocar alguna identificación y la incluyó dentro del rango de entrada, debe marcar el cuadro de Rótulos en la primera fila. En la parte referente a opciones de salida, debe indicar donde desea que Excel coloque los resultados del análisis de datos: Rango de Salida (si es dentro de la misma planilla donde tiene los datos), en una hoja nueva (otra planilla dentro del mismo archivo) o en un libro nuevo (archivo nuevo). Luego se seleccionará el cuadro RESUMEN DE ESTADÍSTICAS, y llegaremos a los siguientes resultados para la Floresta A:

Columna1	
Media	35.05
Error típico	2.875981138
Mediana	27
Moda	16
Desviación estándar	31.50479489
Varianza de la muestra	992.5521008
Curtosis	3.849254513
Coeficiente de asimetría	1.876004202
Rango	156
Mínimo	5
Máximo	161
Suma	4206
Cuenta	120

d) Aplicando el anterior procedimiento para la floresta B, obtendremos los siguientes resultados:

Columna1	
Media	32.2
Error típico	0.898769779
Mediana	32
Moda	30
Desviación estándar	9.845529634
Varianza de la muestra	96.93445378
Curtosis	0.341914773
Coeficiente de asimetría	0.005765898
Rango	54
Mínimo	7
Máximo	61
Suma	3864
Cuenta	120

1.5 Uso de la herramienta ANÁLISIS DE DATOS para construir HISTOGRAMAS

Se utilizaran los datos referentes al levantamiento forestal descrito en el inciso 1.4, para esto, se deben recordar algunos pasos descritos en el inciso 1.2.2.1, con la salvedad de que para la construcción directa de los histogramas, a la par de la columna de datos referentes a la variable a analizar, debe colocarse una columna con los límites superiores de cada una de las clases clase, previamente establecidas.

Se tomará para este ejemplo, los datos de la floresta A.

a) Seleccione en el menú HERRAMIENTAS la opción ANÁLISIS DE DATOS y luego dentro de ésta: HISTOGRAMA.

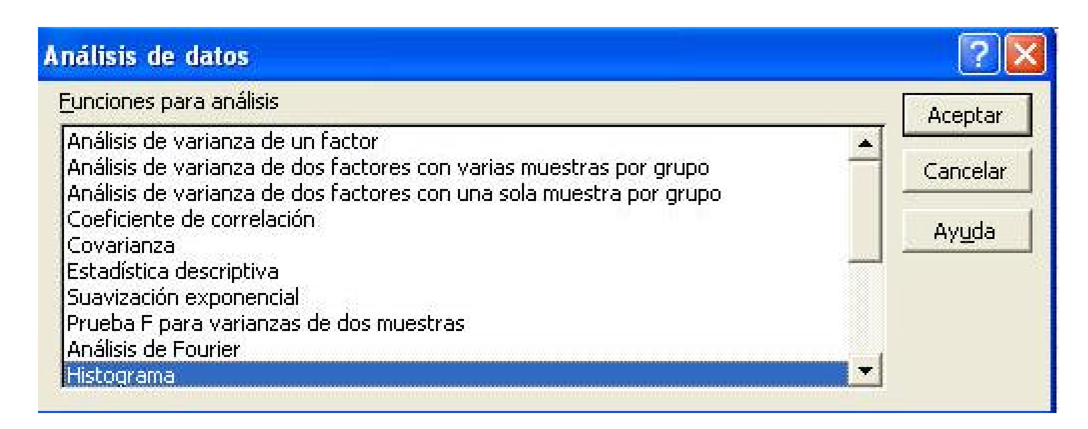


Figura 41 Función HISTOGRAMA

b) Luego de presionar ACEPTAR en la Figura 42, se desplegará la siguiente pantalla:

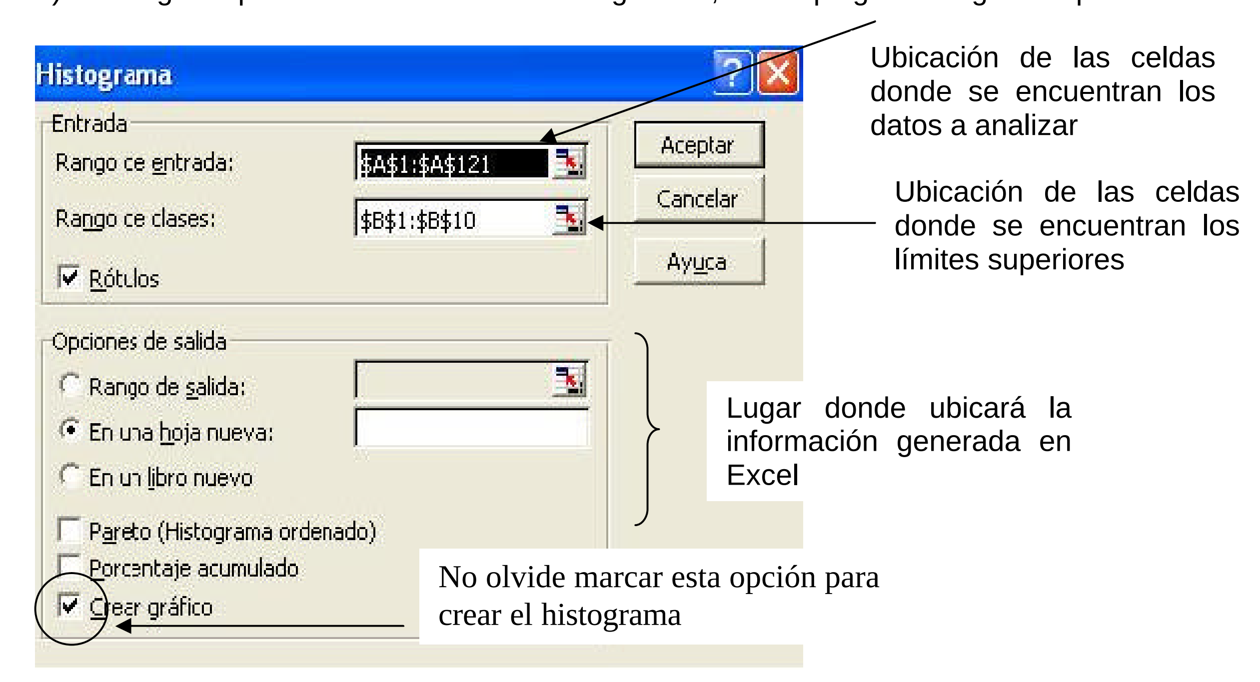
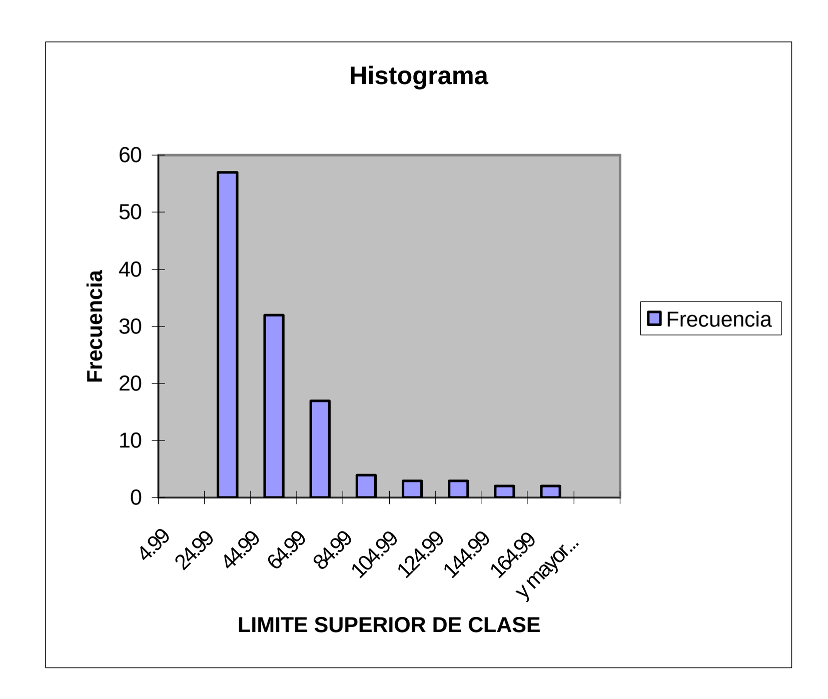


Figura 42 Opciones para la construcción de un HISTOGRAMA

c) Al dar clic en ACEPTAR se generará el siguiente gráfico



LIMITE SUPERIOR DE CLASE	*Frecuencia*
4.99	0
24.99	57
44.99	32
64.99	17
84.99	4
104.99	3
124.99	3
144.99	2
164.99	2
y mayor...	0

Figura 43 Histograma para los datos de diámetro de los árboles de la Floresta A y tabla de frecuencia para cada clase

d) Observe en la Figura 43, que Excel construyó un gráfico de barras, pero haciendo algunas correcciones, se logrará construir un histogramas. Vea los siguientes pasos:

d.1 De un clic con el botón del lado derecho (o el izquierdo) del *mouse* sobre cualquiera de las barras y observará lo siguiente

= =SERIES("Frecuencia",Hoja4!$A$2:$A$11,Hoja4!$B$2:$B$11,1)

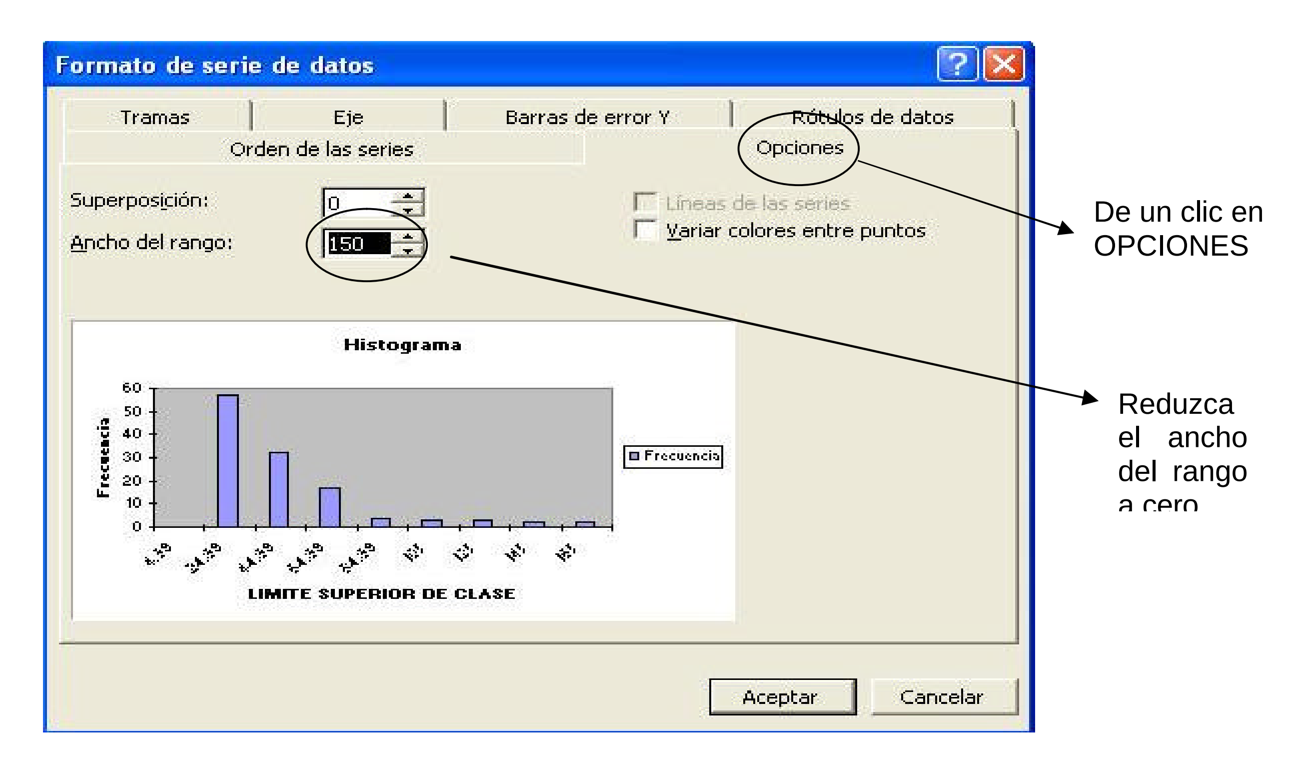
Note que en la celda A11 y B11 corresponden al límite que Excel denomina y mayor......, el cual no debe aparecer en el gráfico, por lo tanto sustitúyalos por A10 y B10.

d.2 De un clic con el botón del lado derecho sobre cualquiera de las barras del gráfico que generó Excel y se desplegará un cuadro similar al que se muestra en la Figura 44. Luego seleccione FORMATO DE SERIE DE DATOS



Figura 44 Opción FORMATO DE SERIES DE DATOS

Luego de dar un clic en FORMATO DE SERIE DE DATOS, se desplegará el siguiente cuadro, realice los cambios que en él se muestran y vea la forma final del Histograma en la Figura 45





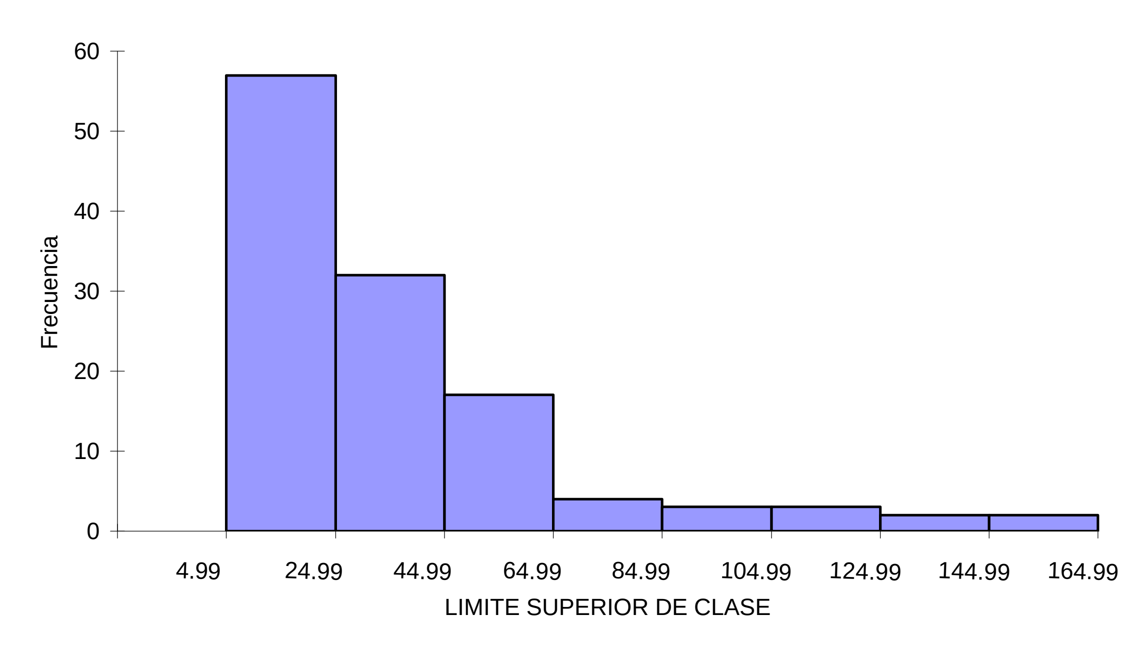


Figura 45 HISTOGRAMA (versión final)

1.6 Uso de EXCEL para construir un diagrama de cajas y alambres (BOX PLOT)

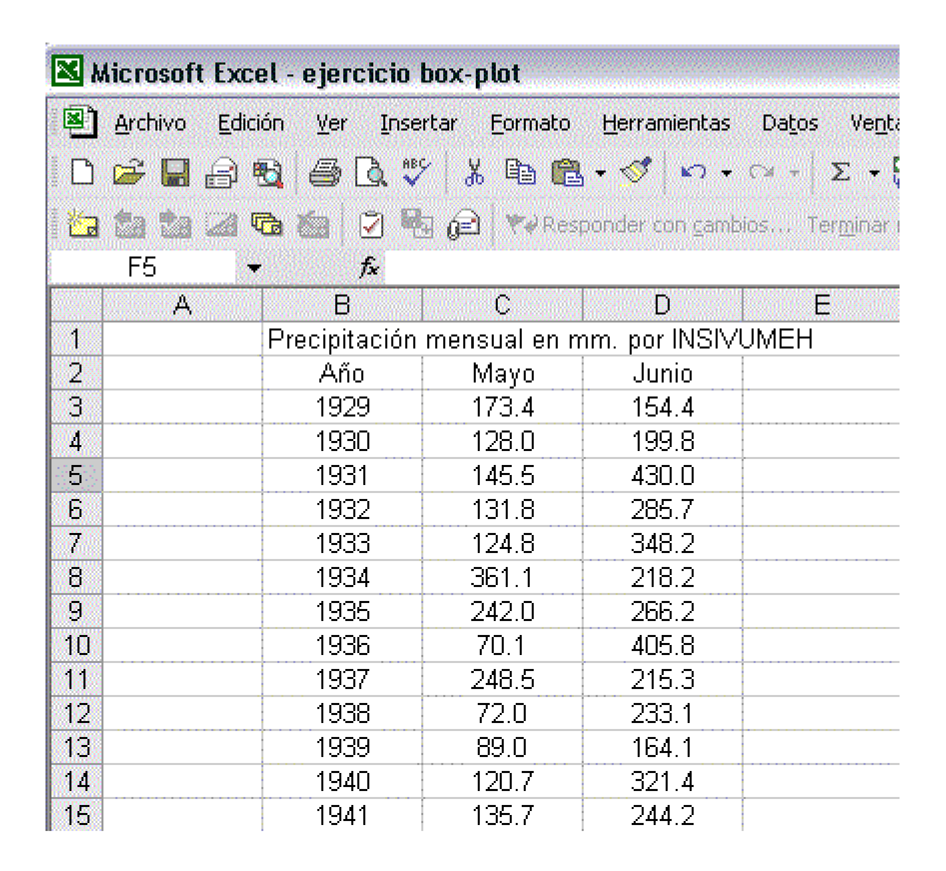
Para construir un box plot (caja de dispersión) necesitamos determinar el valor del primer y el tercer cuartil, el valor del a mediana, y los valores mínimo y máximo de la variable analizada.

Todos estos estadísticos son provistos por la opción de estadística descriptiva del menú de Análisis de Datos de Excel, con excepción del primer y tercer cuartil.

En la carpeta *Mis documentos* o *My documents* dependiendo si su sistema operativo está en inglés o en español podrá localizar el archivo que contiene los datos para el presente ejercicio. La ruta sería *C:\Mis documentos\Estadistica\ejercicio box plot.xls* o bien *C:\My documents\Estadistica\ejercicio box plot.xls*.

Abra el archivo *ejercicio box plot.xls* y vea que la información se encuentra localizada en columnas. Se trata de los registros de precipitaciones pluviales para los meses de mayo y junio de los años 1929 a 1996. El propósito es hacer una comparación gráfica de la dispersión de la precipitación pluvial para estos dos meses en el período mencionado.

El archivo con el presente ejercicio también puede ser descargado en http://www.byrong.tk en el área de publicaciones y apartado de estadística. El archivo luce de la siguiente manera:



	A	B	C	D	E
1		Precipitación mensual en mm. por INSIVUMEH			
2		Año	Mayo	Junio	
3		1929	173.4	154.4	
4		1930	128.0	199.8	
5		1931	145.5	430.0	
6		1932	131.8	285.7	
7		1933	124.8	348.2	
8		1934	361.1	218.2	
9		1935	242.0	266.2	
10		1936	70.1	405.8	
11		1937	248.5	215.3	
12		1938	72.0	233.1	
13		1939	89.0	164.1	
14		1940	120.7	321.4	
15		1941	135.7	244.2	

Procedimiento general para la construcción de Box Plot en Excel

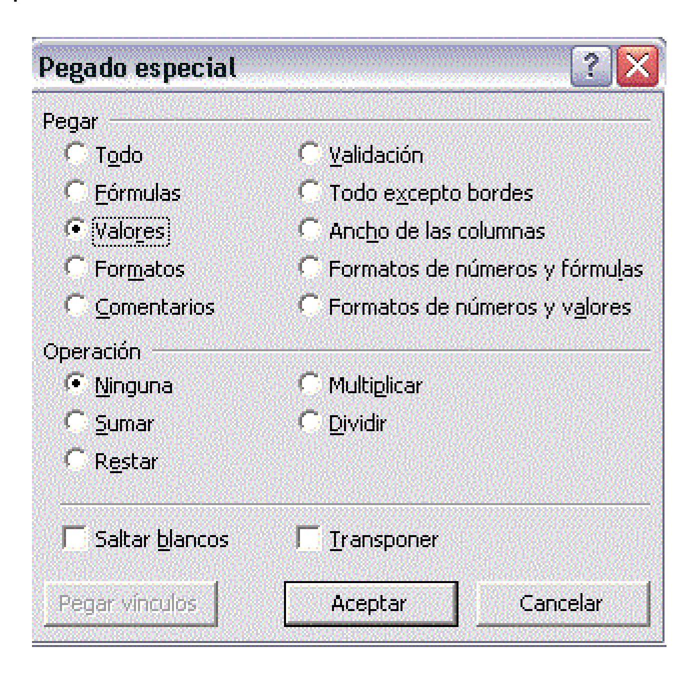
a. En la celda F5 escriba *Precipitación*. En la celda F6 escriba 1er. Cuartil y a su lado en la celda G6 escriba +PERCENTIL(C3:C70,0.25). En la celda F7 escriba *mínimo*, y a su lado en la celda G7 escriba =+MIN(C3:C70).

   En la celda F8 escriba *Mediana* y a su lado en la celda G8 escriba =+MEDIANA(C3:C70). En la celda F9 escriba *máximo* y a su lado en la celda G9 escriba =+MAX(C3:C70).

   Para finalizar esta parte en la celda F10 escriba *3er. Cuartil*. A su lado en la celda G10 escriba =+PERCENTIL(C7:C74,0.75). El aspecto en la hoja de cálculo debiera ser el siguiente:

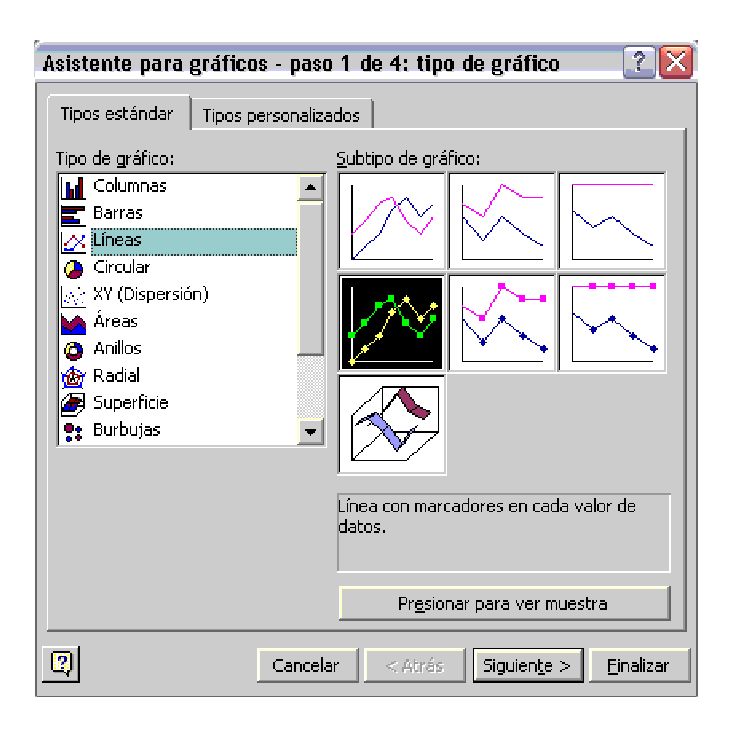
Precipitación	
1er. Cuartil	71.525
mínimo	11.0
mediana	123.4
máximo	361.1
3er. Cuartil	152.2

b. Seleccione el rango de celda F5 a G10 y luego usando el botón derecho del ratón seleccione la alternativa *Copiar*. Manteniendo el rango seleccionado F5 a G10 diríjase al menú principal y elija *Edición/Pegado especial*. En el cuadro de diálogo que aparece seleccione la opción *valores*, como se muestra.

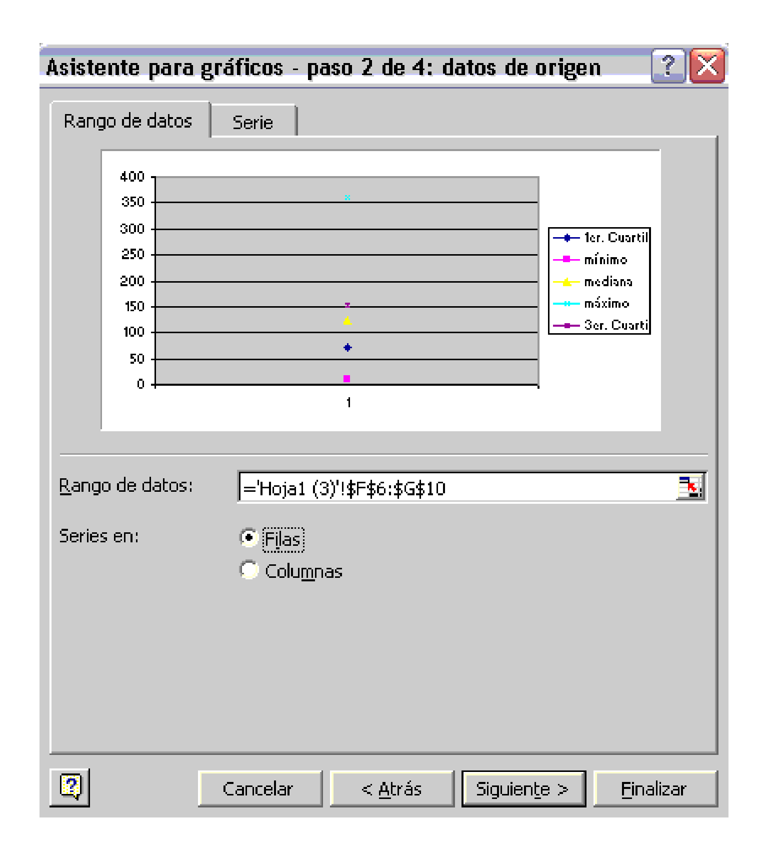


Presione el botón *Aceptar* y verá que aparentemente no se opera cambio alguno. En realidad acaba de convertir las fórmulas introducidas en el paso anterior en valores que pueden usarse para cálculos posteriores**.**

c. Seleccione el rango F6 a G10 y luego del menú principal elija *Insertar/Gráfico*. En tipo de gráfico seleccione *Líneas* y en subtipo de gráfico *Línea con marcadores en cada valor*, como se muestra.

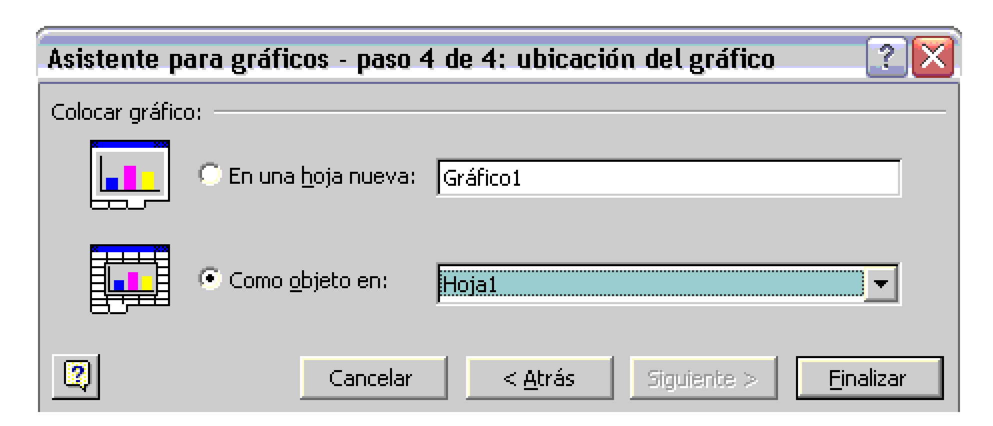


d. Presionar el botón *siguiente* y en la ventana siguiente elegir series en filas como se indica.

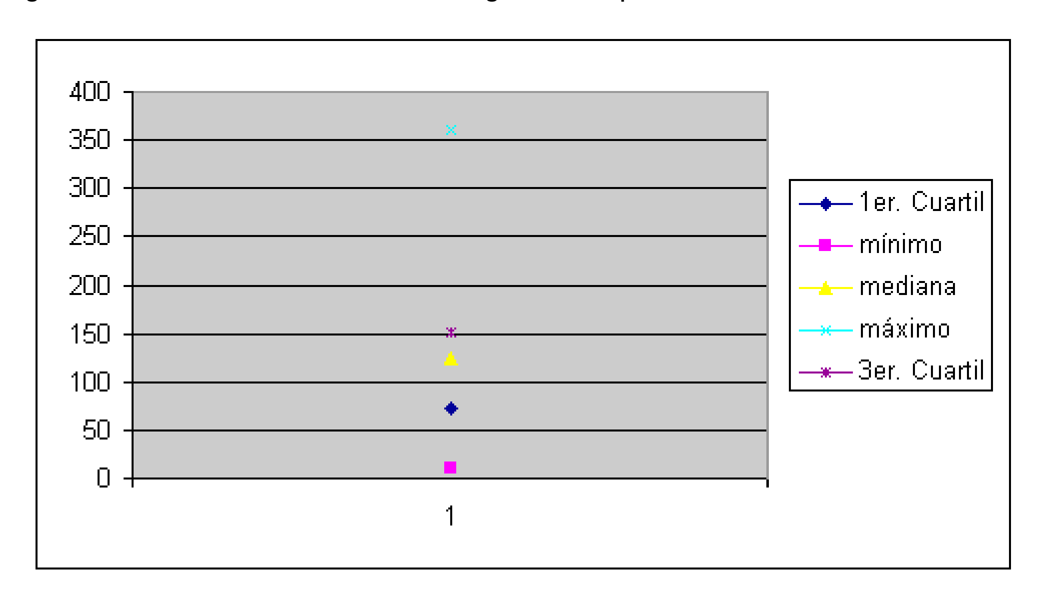


e. Presione el botón *siguiente* y podrá ver la ventana del asistente de gráficos en su paso 3. Si lo desea podrá agregarle un título al gráfico. Nosotros por el momento no le colocaremos distintivo alguno de título.

f. Al presionar nuevamente el botón *siguiente* llegaremos al último paso del asistente de gráficos. El paso 4 permite decidir en dónde se colocará nuestro gráfico. En nuestro caso elegiremos *Como objeto en la hoja1 y Finalizar*, como se muestra a continuación.



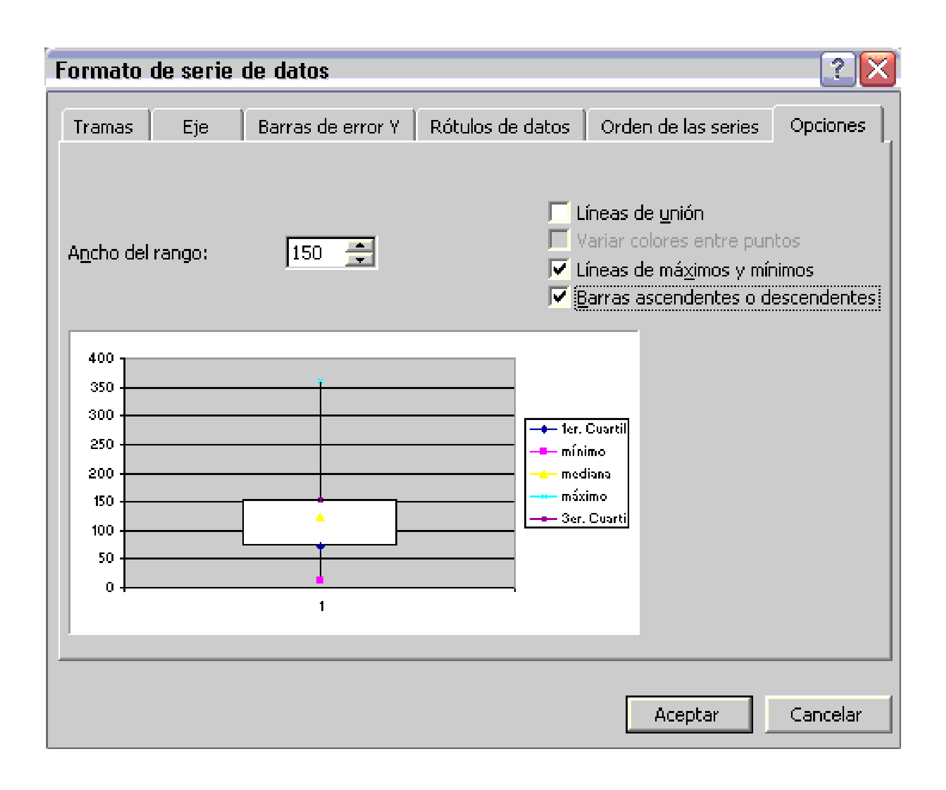
El gráfico resultante deberá tener el siguiente aspecto:



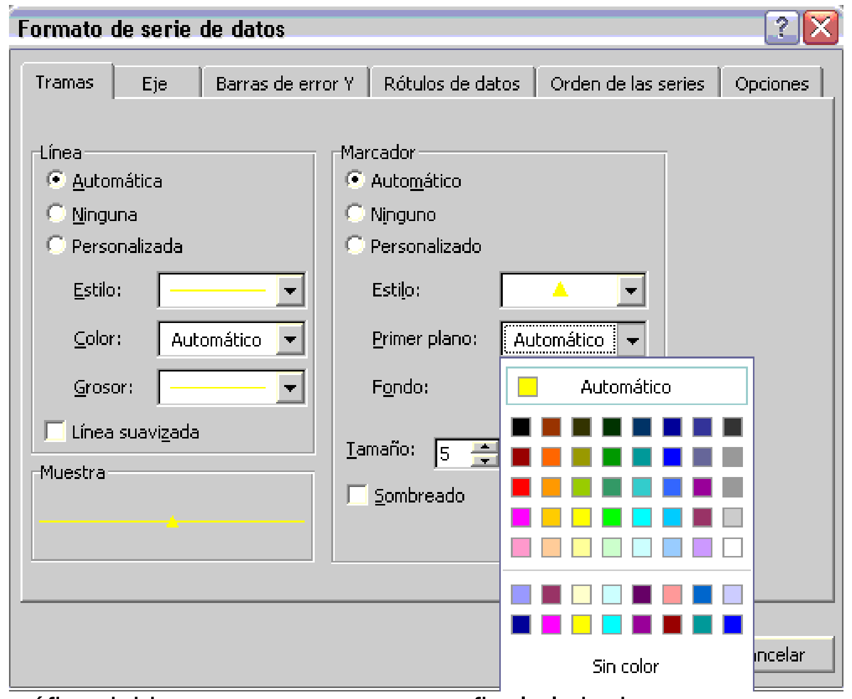
g. Haga clic derecho sobre uno de los puntos del gráfico y escoja del menú resultante la alternativa *Formato de la serie de datos*

Formato de serie de datos...

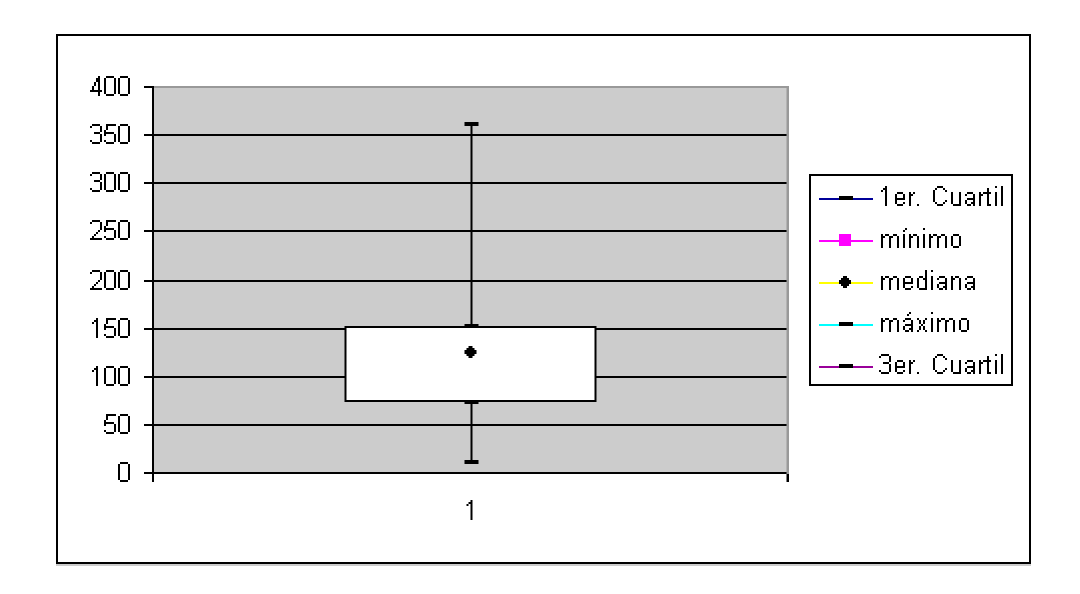
h. En la ventana resultante ir a la pestaña *Opciones* y de las alternativas que incluye seleccionar *Líneas de máximos y mínimos;* y *Barras ascendentes ó descendentes* como se indica.



i. Para finalizar el gráfico y dar un aspecto uniforme en color, nuevamente haga clic derecho sobre el punto central del gráfico y elija la alternativa *Formato de la serie de datos.* Ahora seleccione la pestaña *Tramas*, y en el apartado de *Marcador* cambie el color a negro como se indica.



Nuestro gráfico debiera tener como aspecto final el siguiente:



2. DISTRIBUCIONES DE PROBABILIDAD

2.1 DISTRIBUCIONES DISCRETAS DE PROBABILIDAD

Con Excel se tiene la posibilidad de calcular probabilidades para varias distribuciones discretas, incluyendo la binomial, la de Poisson e hipergeométrica. En esta sección se describirá cómo se puede emplear Excel para calcular las probabilidades en la distribuciones binomial y de Poisson.

a) Distribución binomial

La función DISTR.BINOM. devuelve la probabilidad de una variable aleatoria discreta siguiendo una distribución binomial. Use DISTR.BINOM en problemas con un número fijo de pruebas o ensayos, cuando los resultados de un ensayo son sólo éxito o fracaso, cuando los ensayos son independientes y cuando la probabilidad de éxito es constante durante todo el experimento.

Sintaxis

DISTR.BINOM(núm_éxito;ensayos;prob_éxito;acumulado)

Núm_éxito, es el número de éxitos en los ensayos.

Ensayos, es el número de ensayos independientes.

Prob_éxito, es la probabilidad de éxito en cada ensayo.

Acumulado, es un valor lógico que determina la forma de la función. Si el argumento acumulado es VERDADERO, DISTR.BINOM devuelve la función de distribución acumulada, que es la probabilidad de que exista el máximo número de éxitos deseados; si es FALSO, devuelve la función de masa de probabilidad, que es la probabilidad de que un evento se reproduzca un número de veces igual al argumento núm_éxito.

Ejemplo: Supongamos que tenemos un experimento binomial con *n* = 10 ensayos independientes y *p* = 0.30 (probabilidad de éxito). Vamos a suponer también que nos interesa la probabilidad de *x*=4 éxitos en los 10 intentos. Los pasos que siguen describen cómo usar Excel para obtener la probabilidad binomial deseada.

Paso 1 Seleccione una celda, en la hoja de cálculo donde desee que aparezca la probabilidad binomial.

Paso 2 Seleccione el menú desplegable INSERTAR.

Paso 3 Seleccione la opción FUNCION

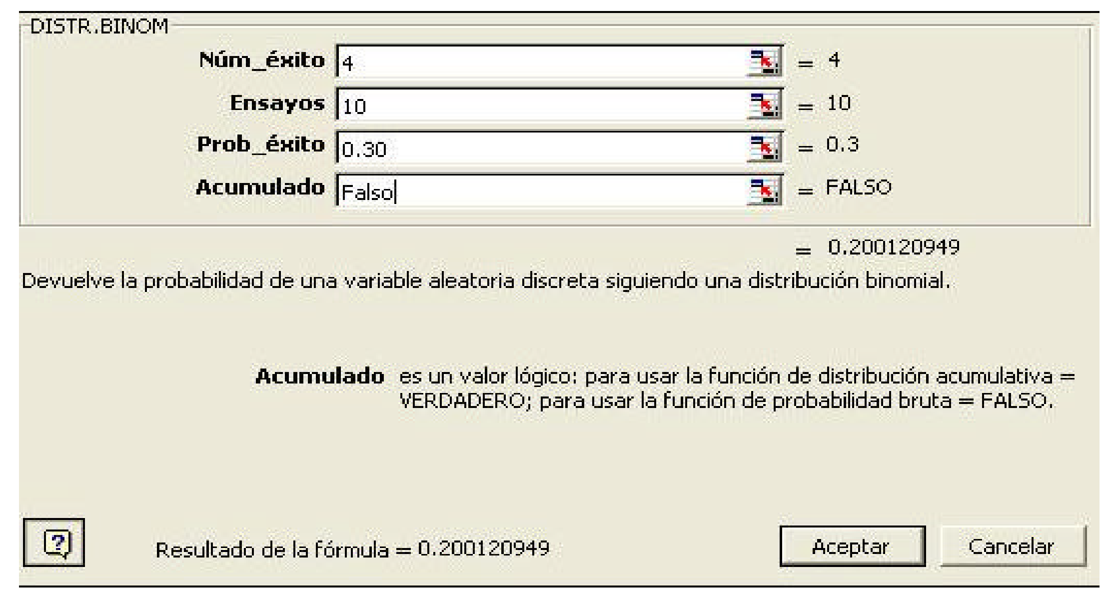
Paso 4 Cuando aparezca el cuadro de diálogo del auxiliar de funciones :

Escoja Estadísticas en el cuadro Categoría de la función.

Escoja DISTR.BINOM. del cuadro de diálogo Nombre de la función

Seleccione Aceptar.

Paso 5 Cuando aparezca el cuadro de diálogo del auxiliar de funciones, teclee:



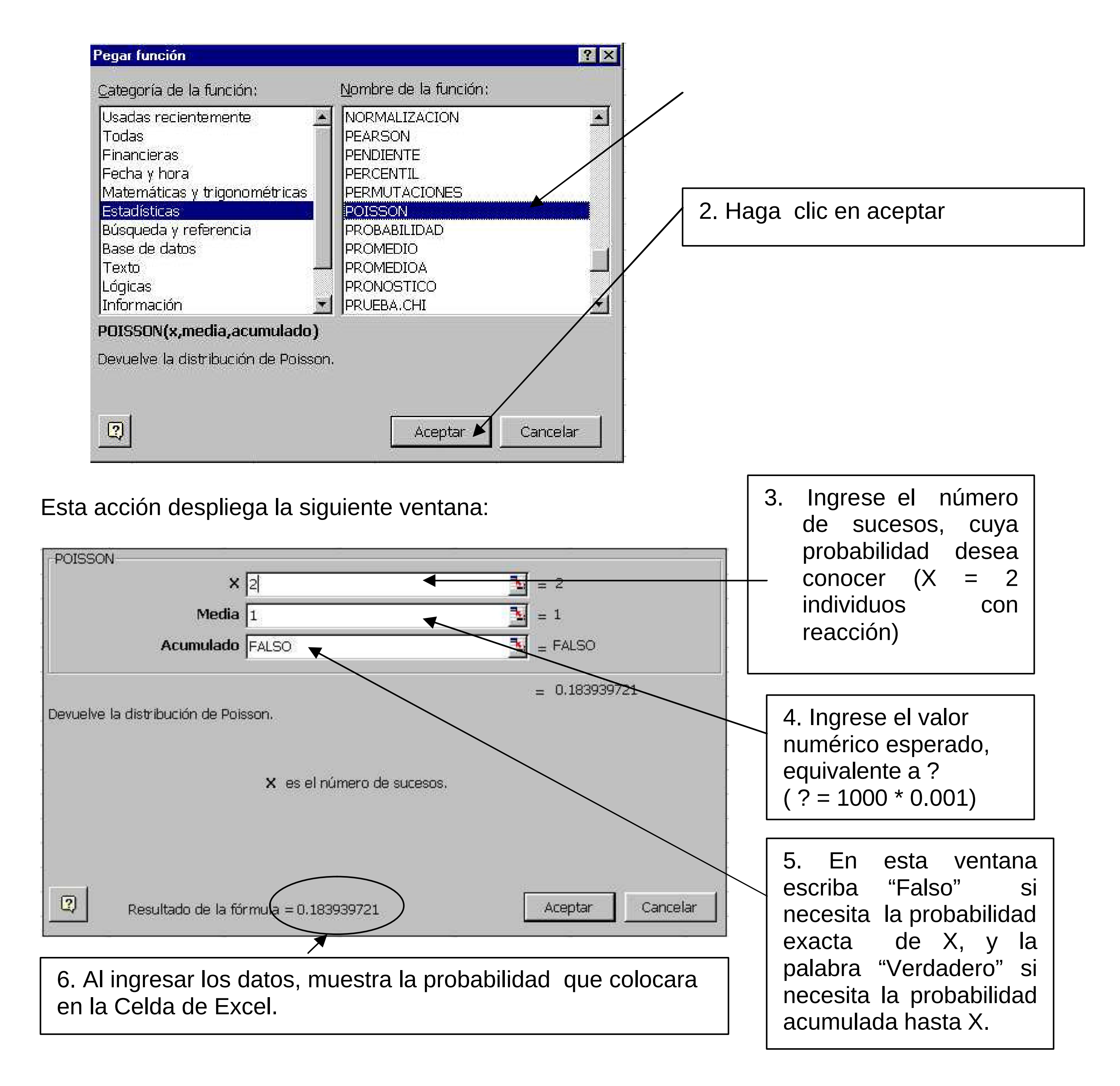
Si se hubiese presionado “Verdadero” en la parte referente a ACUMULADO, el valor obtenido correspondería a la probabilidad de obtener un máximo de 4 éxitos.

b) Distribución de Poisson

Ejemplo:
Si la probabilidad de que un individuo sufra una reacción por la inyección de un determinado suero es de 0.001, determine la probabilidad de que de un total de 1000 individuos se encuentre exactamente 2 individuos con reacción. RESPUESTA: 0.1839

METODOLOGÍA

1. Haga un clic en el botón *fx*, lo cual desplegará la siguiente ventana donde debe seleccionar “POISSON”



2.2 DISTRIBUCIONES CONTINUAS DE PROBABILIDAD

Excel tiene posibilidad de calcular valores de probabilidad de diversas distribuciones continuas, incluyendo la normal y la exponencial. En esta sección se describirá cómo se utiliza Excel para calcular probabilidades de cualquier distribución normal. Los procedimientos para las distribuciones exponencial y otras continuas son similares a los que se describirán.

Ejemplo:

La empresa Grear Tire Company acaba de desarrollar un neumático radial con banda de acero que venderá a través de una cadena nacional de tiendas de descuento. Como ese neumático es producto nuevo, la dirección de Grear cree que la garantía de millas recorridas que se ofrece con el neumático será un factor importante en la aceptación. Antes de formalizar esa política, la dirección desea contar con información acerca de las millas que duran los neumáticos.

En pruebas reales de carretera, el grupo de ingeniería de Grear ha estimado que el promedio de distancia recorrida es μ = 36,500 millas y la desviación estándar es σ = 5,000. Además, los datos reunidos indican que la variable en estudio sigue aproximadamente una distribución normal. ¿Qué porcentaje de neumáticos se puede esperar que duren más de 40,000 millas?

Solución: Los pasos siguientes describen cómo utilizar Excel para obtener la probabilidad normal.

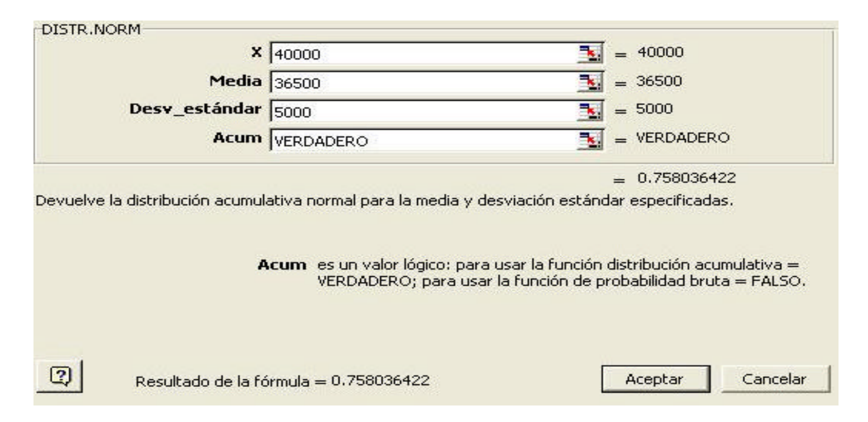
Paso 1 Seleccionar una celda en la hoja de trabajo donde desee que aparezca el valor de probabilidad normal.

Paso 2 Seleccionar el menú desplegable INSERTAR

Paso 3 Seleccionar la opción FUNCION

Paso 4 Cuando aparece el cuadro de diálogo PEGAR FUNCION:
Seleccionar Estadísticas del cuadro Categoría de la función
Seleccionar DISTR.NORM del cuadro Nombre de la función
Seleccionar Aceptar

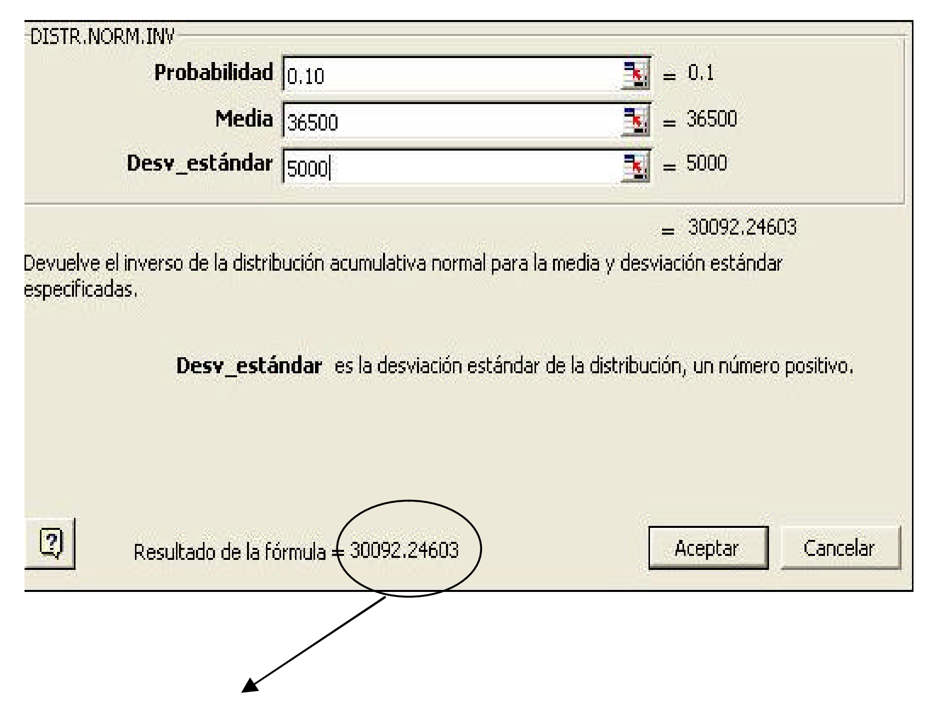
Paso 5 Cuando aparezca el cuadro de diálogo DISTR.NORM, teclee lo siguiente:



Aparecerá 0.758 en la cela que seleccionó en el paso 1, indicando que la probabilidad de que la duración en millas sea menor que o igual a 40,000 millas es de 0.758. La probabilidad de que la duración sea mayor que 40,000 millas es de 1– 0.758 = 0.242

Excel también utiliza cálculos inversos para convertir determinada probabilidad normal acumulada en un valor de la variable aleatoria. Por ejemplo, ¿qué duración en millas debe ofrecer Grear, si la empresa no desea que más del

10% de sus neumáticos califiquen para la garantía? Para calcular la garantía con Excelse sigue el procedimiento que se acaba de describir. Sin embargo, hay necesidad de dos cambios: en el paso 4 se elige DISTR.NORM.INV en el cuadro Nombre de la función; en el paso 5 teclee lo siguiente:



La garantía de duración de los neumáticos es de aproximadamente 30,100 millas

Función DISTR.NORM.ESTAND

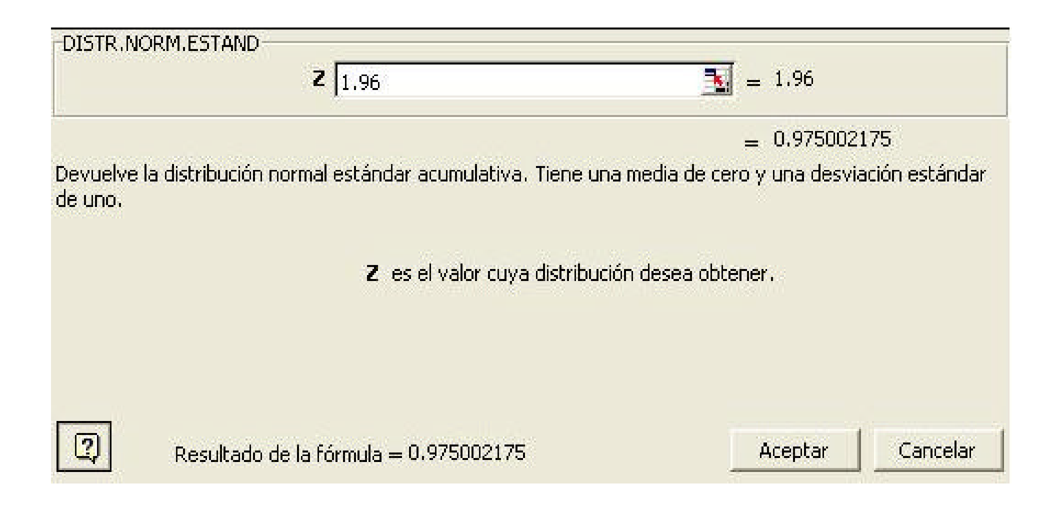
Devuelve la función de distribución normal estándar acumulativa. La distribución tiene una media de 0 (cero) y una desviación estándar de uno. Use esta función en lugar de una tabla estándar de áreas de curvas normales.

Sintaxis

DISTR.NORM.ESTAND(z)Z es el valor cuya distribución desea obtener.

Ejemplo:

Calcule el valor de probabilidad acumulada para un valor de z =1.96



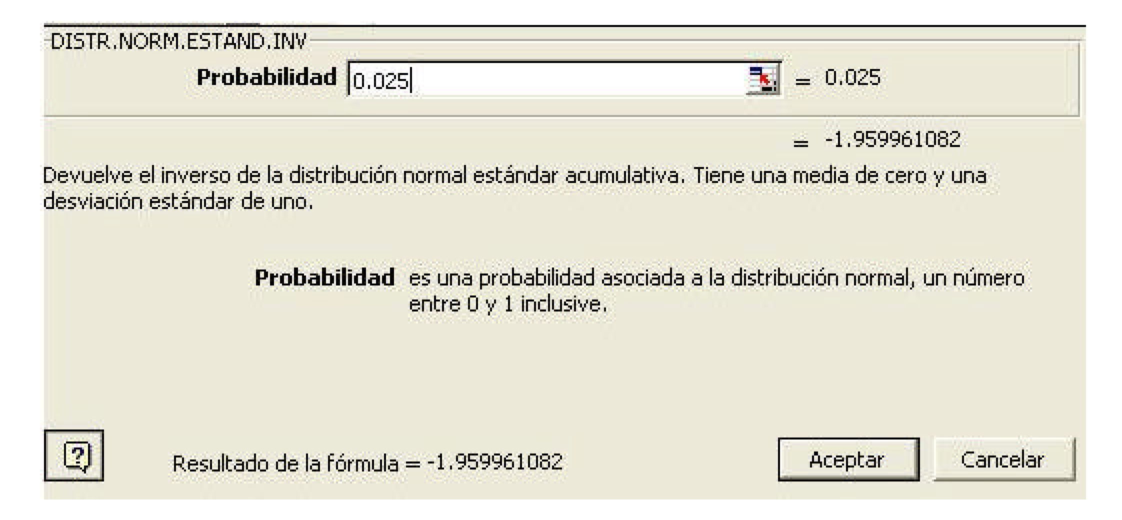
Función DISTR.NORM.ESTAND.INV

Devuelve el inverso de la distribución normal estándar acumulativa. La distribución tiene una media de cero y una desviación estándar de uno.

Sintaxis
DISTR.NORM.ESTAND.INV(probabilidad)

Probabilidad, es una probabilidad que corresponde a la distribución normal.

Ejemplo
Calcule el valor de z que corresponde a un valor de probabilidad de 0.025



El valor de z correspondiente a un valor de probabilidad de 0.025 es de aproximadamente –1.96

3. PRUEBA DE HIPÓTESIS

3.1 Prueba de Diferencias Pareadas: Ejemplo de aplicación -Introducción a la Probabilidad y Estadística, William Mendenhal-

Un fabricante deseaba comparar la resistencia al desgaste de dos tipos distintos de neumáticos A y B. Para hacer la comparación, se asignó al azar un neumático del tipo A y uno del tipo B a las ruedas posteriores de 20 automóviles.

Los autos recorrieron un número específico de kilómetros y se observó el desgaste de cada neumático. Estos valores aparecen en la tabla 1.

Presentan los datos suficiente evidencia para concluir que hay diferencia en el desgaste promedio de los dos tipos de neumáticos?

Tabla 1:

Automóvil	1	2	3	4	5	6	7	8	9	10
Neumático A	10.6	9.8	12.3	9.7	8.8	10	9.9	9	12.1	8.9
Neumático B	10.2	9.4	11.8	9.1	8.3	10.1	9.2	11.2	11	8.2

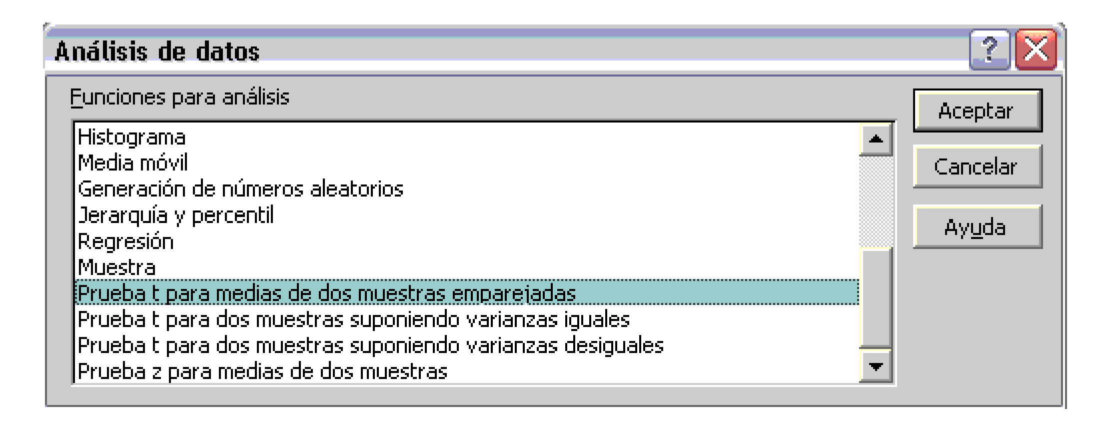
Automóvil	11	12	13	14	15	16	17	18	19	20
Neumático A	10.1	11	11.8	9.9	12.2	12.3	10.5	8.8	8.6	9.2
Neumático B	10.1	10	10.3	10.4	11.1	11.3	9.3	8.5	10.3	11

3.1.1 Procedimiento general de resolución empleando Excel

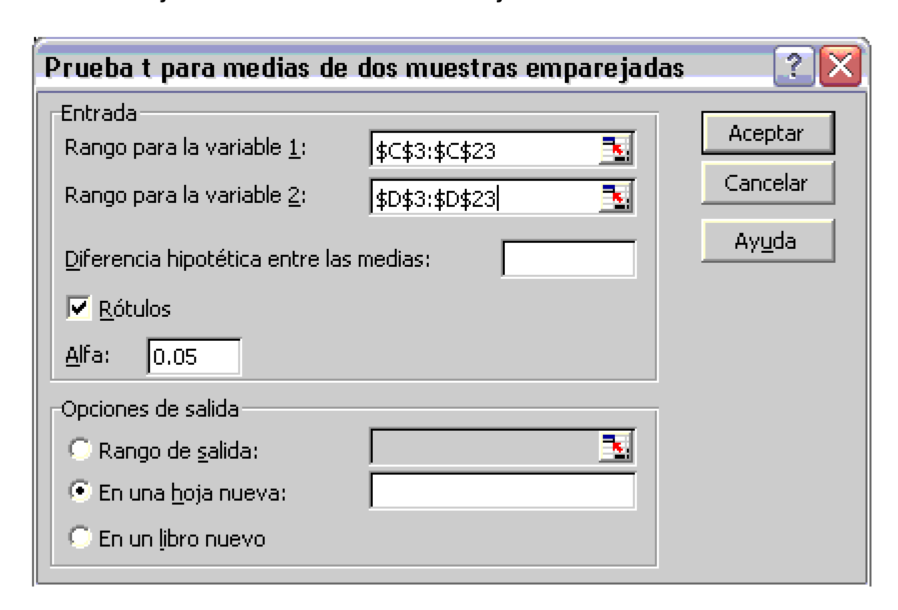
En la carpeta Mis documentos ó My documents dependiendo si su sistema operativo está en inglés o en español podrá localizar el archivo que contiene los datos para el presente ejercicio. La ruta sería C:\Mis documentos\Estadistica\ejercicio pruebas de medias pareadas.xls o bien C:\My documents\Estadistica\ejercicio pruebas de medias pareadas.xls.

El archivo con el presente ejercicio también puede ser descargado en http://www.byrong.tk en el área de publicaciones y apartado de estadística.

a. Seleccione del menú principal Herramientas/Análisis de datos. Una vez aparezca en pantalla el cuadro de dialogo elija Prueba t para medias de dos muestras emparejadas, como se indica.



b. Presione luego el botón *Aceptar*. En *rango para la variable 1* escribir C3:C23. En *rango para la variable 2* escribir D3:D23. No olvide marcar la casilla de verificación *Rótulos* dado que las variables se encuentran rotuladas. Elija si desea ver los resultados en la hoja actual o bien en una hoja nueva.



c. Luego de presionar el botón *Aceptar* aparecen en una hoja nueva (de acuerdo a la elección del cuadro de diálogo anterior) los resultados de la prueba de t. Dependiendo de la hipótesis planteada se recurrirá a ver los valores de probabilidad para una o para dos colas. En el presente caso se revisa la probabilidad para dos colas que resultó ser de 0.3241. También si se compara el valor de t calculada=1.012 que resulta ser menor al valor tabular de 2.093 (t tabulada). Este resultado nos permite concluir que en promedio los dos tipos de neumáticos sufren igual desgaste.

	A	B	C
1	Prueba t para medias de dos muestras emparejadas		
2			
3		*Neumático A*	*Neumático B*
4	Media	10.275	10.04
5	Varianza	1.6325	1.08568421
6	Observaciones	20	20
7	Coeficiente de correlación de Pearson	0.61593587	
8	Diferencia hipotética de las medias	0	
9	Grados de libertad	19	
10	Estadístico t	1.012130183	
11	P(T<=t) una cola	0.16209616	
12	Valor crítico de t (una cola)	1.729131327	
13	P(T<=t) dos colas	0.32419232	
14	Valor crítico de t (dos colas)	2.093024705	

3.2 Pruebas de medias independientes: Ejemplo de aplicación -Introducción a la Probabilidad y Estadística, William Mendenhal-

Una operación de ensamblaje de una planta industrial requiere que un empleado nuevo se someta a un período de entrenamiento de aproximadamente un mes para alcanzar su máxima eficacia.

Se sugirió un nuevo método de entrenamiento y se llevó a cabo de una prueba para comparar el método nuevo con el procedimiento estándar. Dos grupos de nueve empleados nuevos se entrenaron durante un período de tres semanas, un grupo usando el nuevo método y el otro siguiendo el procedimiento de entrenamiento estándar. Al final del período de tres semanas se observó el tiempo en minutos que le tomó a cada empleado ensamblar el dispositivo. Los resultados aparecen en la tabla2.

Presentan los datos suficiente evidencia que indique que el tiempo medio de ensamblaje al final del período de entrenamiento de tres semanas es menor para el nuevo método?

Tabla 2:

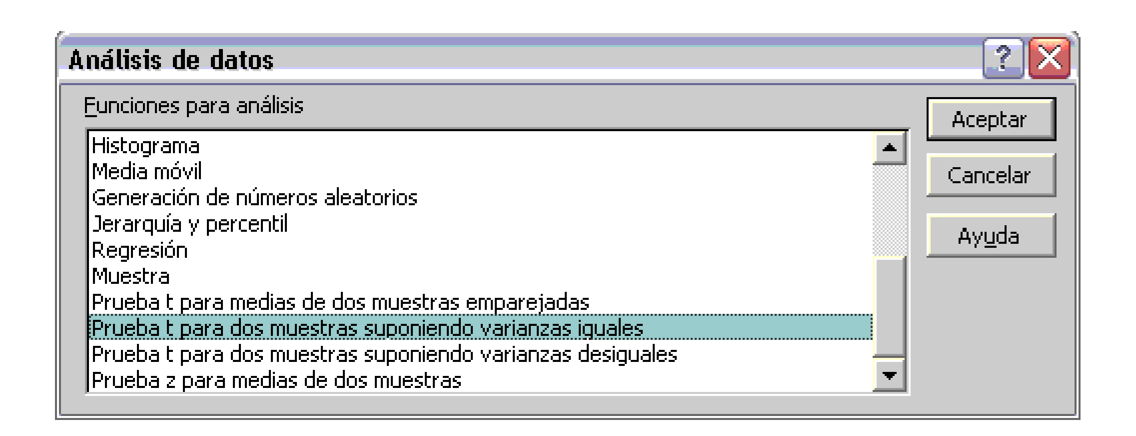
Procedimiento estándar	Procedimiento nuevo
32	35
37	31
35	29
28	25
41	34
44	40
35	27
31	32
34	31

3.2.1 Procedimiento general de resolución empleando Excel

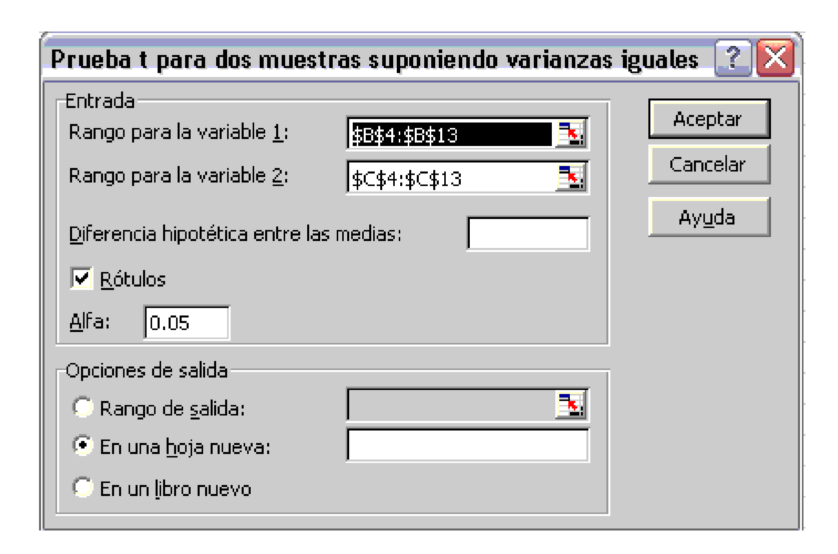
En la carpeta *Mis documentos* ó *My documents* dependiendo si su sistema operativo está en inglés o en español podrá localizar el archivo que contiene los datos para el presente ejercicio. La ruta sería *C:\Mis documentos\Estadistica\ejercicio pruebas de medias independientes.xls* o bien *C:\My documents\Estadistica\ejercicio pruebas de medias independientes.xls*.

El archivo con el presente ejercicio también puede ser descargado en http://www.byrong.tk en el área de publicaciones y apartado de estadística.

a. Seleccione del menú principal *Herramientas/Análisis de datos*. Una vez aparezca en pantalla el cuadro de dialogo elija *Prueba t para dos muestras suponiendo varianzas iguales,* como se indica.



b. Presione luego el botón *Aceptar*. En *rango para la variable 1* escribir B4:B13. En *rango para la variable 2* escribir D4:D13. No olvide marcar la casilla de verificación *Rótulos* dado que las variables se encuentran rotuladas. Elija si desea ver los resultados en la hoja actual o bien en una hoja nueva. También debe resaltarse el supuesto de que las varianzas de ambas muestras son iguales. De no ser cierta esta condición deberá inclinarse por el procedimiento que contempla la desigualdad entre ellas.



c. Luego de presionar el botón *Aceptar* aparecen en una hoja nueva (de acuerdo a la elección del cuadro de diálogo anterior) los resultados de la prueba de t. Dependiendo de la hipótesis planteada se recurrirá a ver los valores de probabilidad para una o para dos colas. En el presente caso se revisa la probabilidad para una cola dado el propósito planteado al inicio de la investigación. La misma resultó ser de 0.059. También si se compara el valor de t calculada=1.649 que resulta ser menor al valor tabular de 1.7458 (t tabulada) la regla de decisión es equivalente. Este resultado nos permite concluir que no hay evidencia de que el tiempo medio de ensamblaje al final del período de entrenamiento para el nuevo método sea menor al método estándar.

	A	B	C
1	Prueba t para dos muestras suponiendo varianzas iguales		
2			
3		*Procedimiento estándar*	*Procedimiento nuevo*
4	Media	35.22222222	31.55555556
5	Varianza	24.44444444	20.02777778
6	Observaciones	9	9
7	Varianza agrupada	22.23611111	
8	Diferencia hipotética de las medias	0	
9	Grados de libertad	16	
10	Estadístico t	1.649484617	
11	P(T<=t) una cola	0.059269899	
12	Valor crítico de t (una cola)	1.745884219	
13	P(T<=t) dos colas	0.118539799	
14	Valor crítico de t (dos colas)	2.119904821	

4. ANÁLISIS DE REGRESIÓN LINEAL SIMPLE

Para ilustrar el análisis de regresión lineal simple usando excel se emplearán los datos de la tabla 3, correspondiente al precio en quetzales de 20 casas.

Tabla 3. Precios y tamaños de 20 casas.

No. Casa	Tamaño (x 100 m^2)	Precio en miles de Q.	No. Casa	Tamaño (x 100 m^2)	Precio en miles de Q.
1	1.8	32	**11**	2.3	44
2	1.0	24	**12**	0.9	19

3	1.7	27	**13**	1.2	25
4	2.8	47	**14**	3.4	50
5	2.2	35	**15**	1.7	30
6	0.8	17	**16**	2.5	43
7	3.6	52	**17**	1.4	27
8	1.1	20	**18**	3.3	50
9	2.0	38	**19**	2.2	37
10	2.6	45	**20**	1.5	28

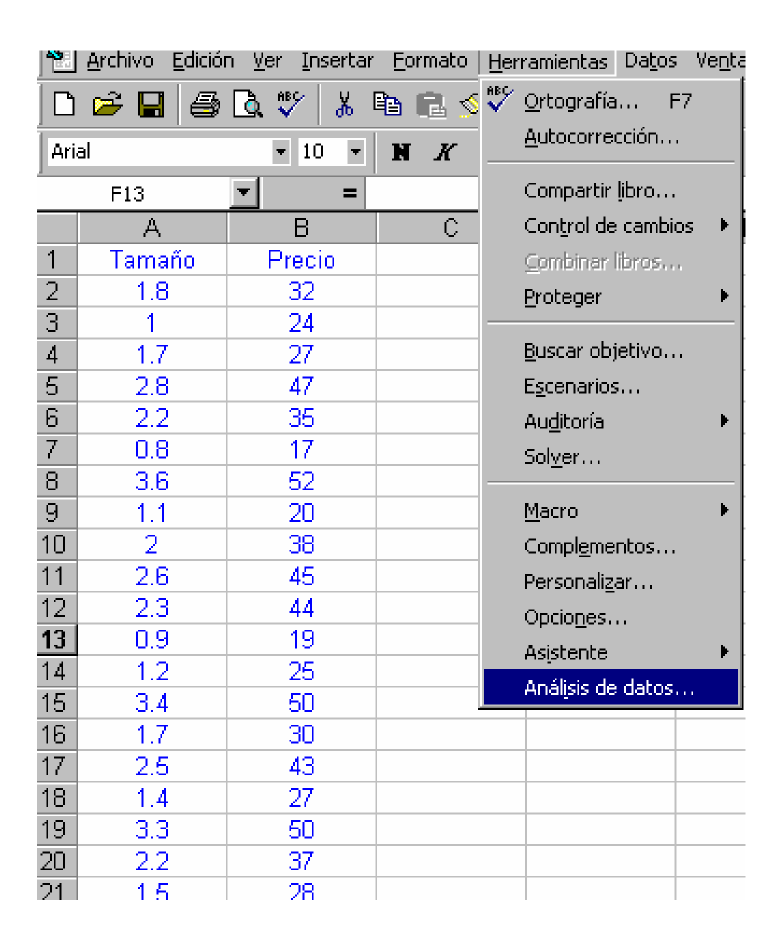
Antes de usar el menú de *Análisis de Datos* debe mencionarse que es necesario realizar un diagrama de dispersión para observar la forma de la relación entre las variables analizadas.

En muchas ocasiones un modelo lineal no presenta un buen ajuste al comportamiento de nuestras variables. En estos casos en aconsejable la prueba de otros modelo, inclusive los no lineales si fuera necesario.

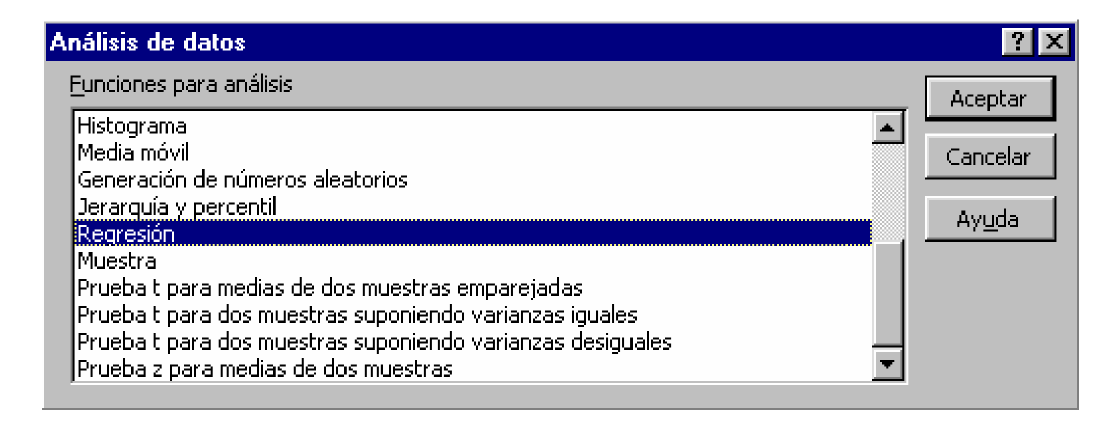
Estos comportamientos pueden ser previstos desde un inicio si se grafica en un plano XY ambas variables analizadas. Este gráfico XY que puede construirse fácilmente con Excel se conoce con el nombre de *Diagrama de dispersión.*

El procedimiento general de análisis de regresión se describe a continuación.

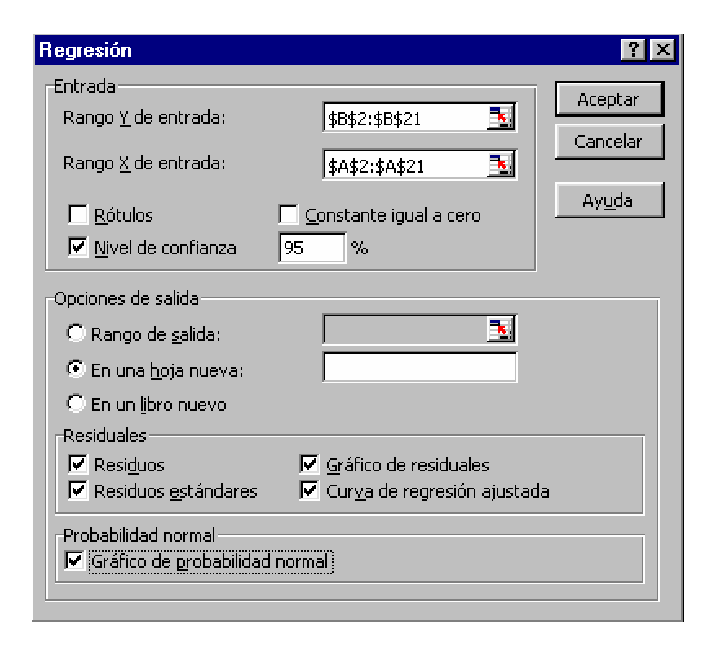
a. Diríjase al menú principal y elija *Herramientas/Análisis de Datos* como se indica a continuación.



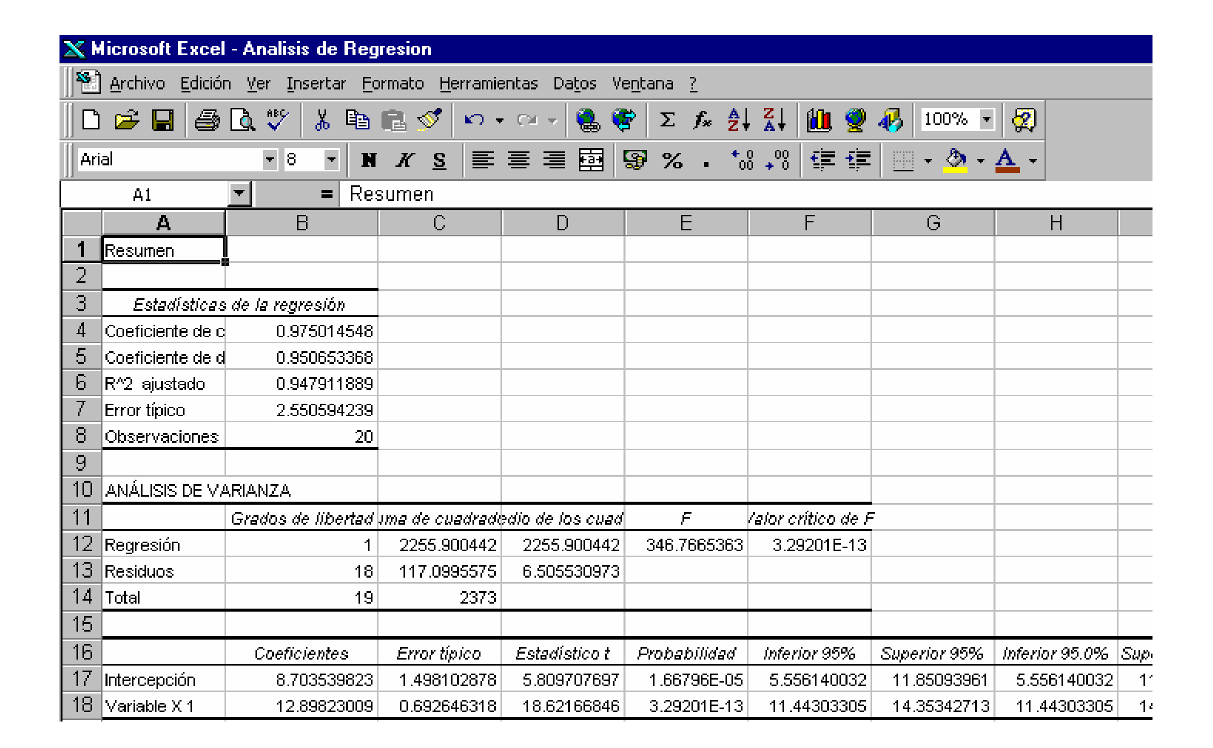
b. La acción anterior desplegará un cuadro de diálogo donde se pregunta por el tipo de análisis a realizar. Escoger *Regresión* y luego *Aceptar*.



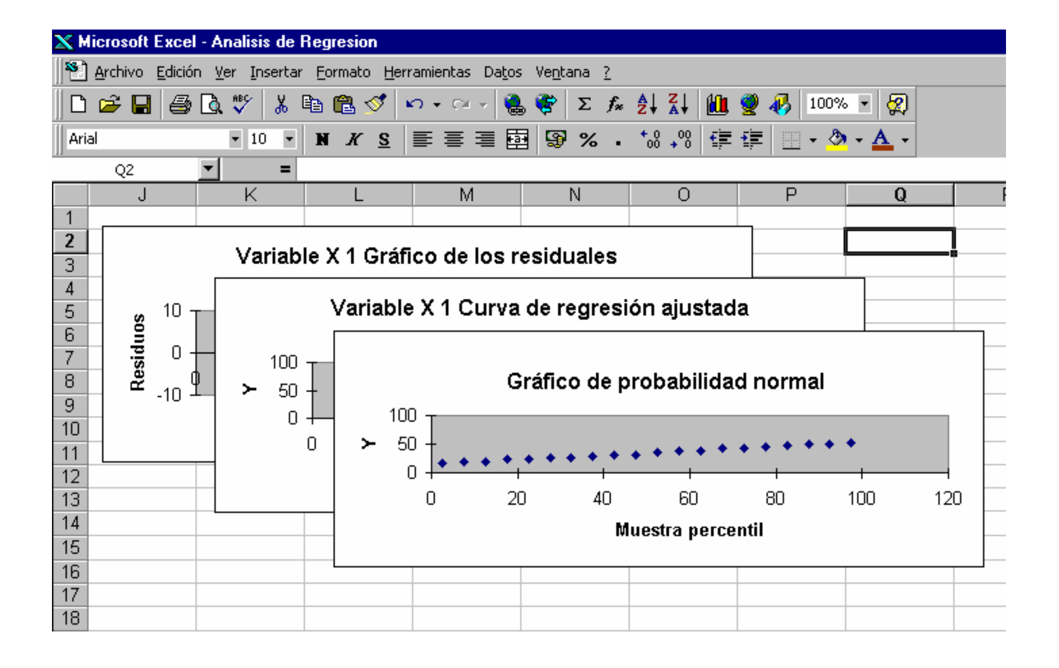
c. El cuadro de diálogo que aparece enseguida presenta varias alternativas, mismas que deberán llenarse de acuerdo a la ilustración.



d. Deberá ingresarse el rango de datos para Y, el rango de datos para X, y solicitar análisis de residuales y también un gráfico de probabilidad normal. Para facilidad de uso posterior se solicita al programa que ubique los resultados en una hoja nueva. Dichos resultados lucirán como los presentados en las siguientes ilustraciones.



	A	B	C	D	E	F	G	H
1	Resumen							
2								
3	*Estadísticas de la regresión*							
4	Coeficiente de c	0.975014548						
5	Coeficiente de d	0.950653368						
6	R^2 ajustado	0.947911889						
7	Error típico	2.550594239						
8	Observaciones	20						
9								
10	ANÁLISIS DE VARIANZA							
11		*Grados de libertad*	*ıma de cuadrados*	*dio de los cuad*	*F*	*/alor crítico de F*		
12	Regresión	1	2255.900442	2255.900442	346.7665363	3.29201E-13		
13	Residuos	18	117.0995575	6.505530973				
14	Total	19	2373					
15								
16		*Coeficientes*	*Error típico*	*Estadístico t*	*Probabilidad*	*Inferior 95%*	*Superior 95%*	*Inferior 95.0%*
17	Intercepción	8.703539823	1.498102878	5.809707697	1.66796E-05	5.556140032	11.85093961	5.556140032
18	Variable X 1	12.89823009	0.692646318	18.62166846	3.29201E-13	11.44303305	14.35342713	11.44303305



5. ANALISIS DE VARIANZA

Ejemplo de aplicación –Estadística: Teoria e Aplicacoes, Levin. D. et al-

El gerente de producción en una fábrica en que están siendo abastecidas cajas con 368 gr. está pretendiendo sustituir una máquina antigua, que afecta directamente el resultado en el proceso de producción. Tres operadores permitirán que el gerente utilice sus equipos por medio de una prueba. Los precios de compra y contrato de servicios para estos tres tipos de máquina son básicamente los mismos.

Para tomar una decisión en cuanto a la compra, el gerente de producción decide hacer una experiencia para saber si existen diferencias significativas entre las tres marcas de máquinas respecto al tiempo medio (en segundos) usado por los trabajadores para completar el proceso.

Quince trabajadores de la fábrica, con experiencia, habilidad, y edades similares son distribuidos aleatoriamente para recibir entrenamiento en una de las tres máquinas, de tal modo que existen cinco trabajadores para cada máquina. Los resultados obtenidos se resumen en la tabla 4.

Presentan los siguientes datos suficiente evidencia para concluir que hay diferencias en el rendimiento medio correspondiente a las cuatro técnicas?

Tabla 4:

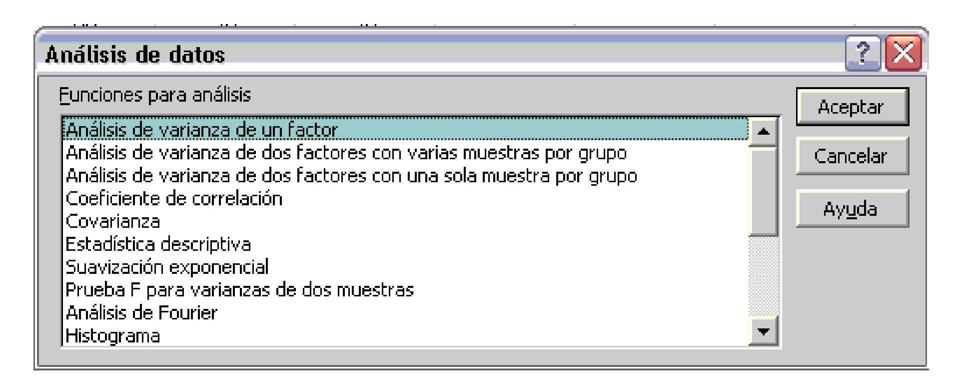
Máquina		
1	2	3
25.4	23.4	20
26.31	21.8	22.2
24.1	23.5	19.75
23.74	22.75	20.60
25.1	21.6	20.40

5.1 Procedimiento general de resolución empleando Excel

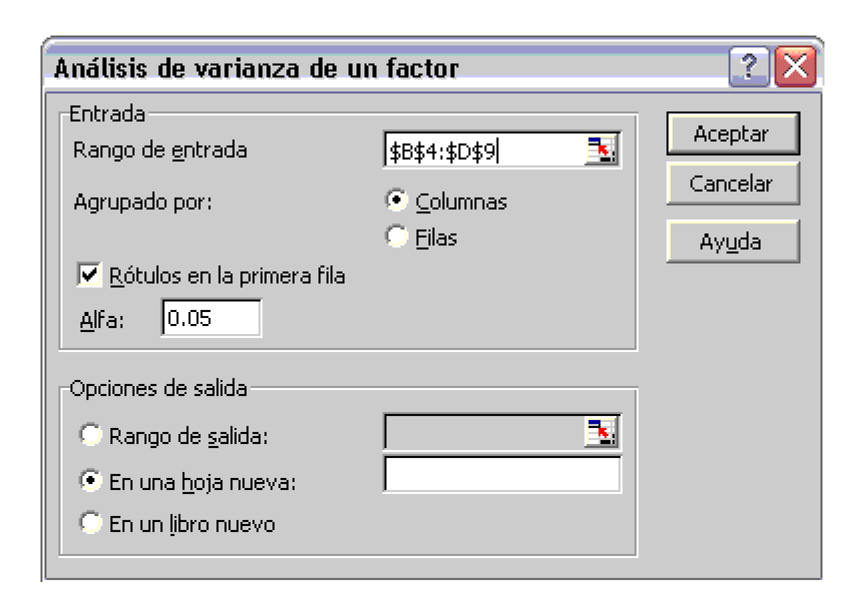
En la carpeta *Mis documentos* ó *My documents* dependiendo si su sistema operativo está en inglés o en español podrá localizar el archivo que contiene los datos para el presente ejercicio. La ruta sería *C:\Mis documentos\Estadistica\ejercicio análisis de varianza de una vía.xls* o bien *C:\My documents\Estadistica\ejercicio análisis de varianza de una vía.xls*.

Al abrir el archivo indicado el aspecto de la pantalla es el siguiente:

a. Seleccione del menú principal *Herramientas/Análisis de datos*. Una vez aparezca en pantalla el cuadro de dialogo elija *Análisis de varianza de un factor,* como se indica.



b. Presione el botón *Aceptar.* En rango de entrada escriba B4:D9. En el apartado de *Agrupado por* seleccione *Columnas*. También seleccione *Rótulos en la primera fila*. Finalmente para conservar los datos originales en una hoja seleccione en *Opciones de salida* la alternativa *en una hoja nueva*.



c. Al presionar el botón *Aceptar* aparecerán en pantalla los resultados correspondientes al análisis de varianza. Destacamos el *p-value* obtenido en el presento ejercicio: 0.0000468 muy inferior al valor crítico de 0.05. Esto nos hace concluir que las tres máquinas evaluadas son diferentes en cuanto al tiempo para completar el proceso de abastecimiento de cereal. Desafortunadamente el menú *Análisis de Datos* de Excel no finaliza el proceso de análisis, al no calcular un procedimiento de comparación múltiple de medias.

Para el efecto usaremos fórmulas de Excel.

	A	B	C	D	E	F	G
1	Análisis de varianza de un factor						
2							
3	RESUMEN						
4	*Grupos*	*Cuenta*	*Suma*	*Promedio*	*Varianza*		
5	Máquina1	5	124.65	24.93	1.0648		
6	Máquina2	5	113.05	22.61	0.778		
7	Máquina3	5	102.95	20.59	0.9205		
8							
9							
10	ANÁLISIS DE VARIANZA						
11	*Origen de las variaciones*	*SC*	*GL*	*CM*	*F*	*Probabilidad*	*Valor crítico para F*
12	Entre grupos	47.164	2	23.582	25.602	4.68404E-05	3.885290312
13	Dentro de los grupos	11.0532	12	0.9211			
14							
15	Total	58.2172	14				

d. El modelo de hoja de trabajo para el cálculo de la prueba múltiple de medias de Tukey se muestra enseguida. Introduzca atentamente las fórmulas que se indican. Se aclara que de todos los valores necesarios para la Prueba de Tukey pueden ser copiados de la tabla resumen de ANDEVA con excepción del estadístico Q de Tukey. Este valor debe ser verificado en la Tabla de Tukey, misma que puede encontrarse fácilmente en cualquier libro de texto de Estadística. En la tabla se ingresa el número de medias a comparar, los grados de libertad del error y el nivel de significancia.

	A	B
1	Comparación múltiple de medias de Tukey	
2		
3	Media del grupo 1	xxx
4	n del grupo 1	x
5	Media del grupo 2	xxx
6	n del grupo 2	x
7	Media del grupo 3	xxx
8	n del grupo 3	x
9	CME (Cuadrado medio del error)	xxx
10	Estadístico Q de Tukey	xxx
11	Comparación del grupo 1 con el grupo 2	
12	Diferencia absoluta	=ABS(B3-B5)
13	Error estándar de la diferencia	=RAIZ((B9/2)*((1/B4)+(1/B6)))
14	Amplitud crítica (W)	=B10*B13
15	Medias del grupo 1 y 2 son	=SI(B12>B14,"diferente";"no diferente")
16	Comparación del grupo 1 con el grupo 3	
17	Diferencia absoluta	=ABS(B3-B7)
18	Error estándar de la diferencia	=RAIZ((B9/2)*((1/B4)+(1/B8)))
19	Amplitud crítica (W)	=B10*B18
20	Medias del grupo 1 y 3 son	=SI(B17>B19,"diferente";"no diferente")
21	Comparación del grupo 2 con el grupo 3	
22	Diferencia absoluta	=ABS(B5-B7)
23	Error estándar de la diferencia	=RAIZ((B9/2)*((1/B6)+(1/B8)))
24	Amplitud crítica (W)	=B10*B23
25	Medias del grupo 2 y 3 son	=SI(B22>B24,"diferente";"no diferente")

e. Luego de introducir los valores correspondientes en la hoja modelo anterior tendremos los resultados para la prueba de Tukey. Para el presente caso las tres medias resultan ser diferentes entre sí. De ellas, la máquina 3 es la que presenta el menor tiempo para completar el proceso de abastecimiento de cereal, por lo que se recomienda su uso.